1-1-1976

A consideration of the generalizability of student ratings of college teachers and courses.

Harold Olaf Bettencourt
University of Massachusetts Amherst

A CONSIDERATION OF THE GENERALIZABILITY OF STUDENT RATINGS OF COLLEGE TEACHERS AND COURSES

A Dissertation Presented

by

HAROLD OLAF BETTENCOURT

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

June 1976

Department of Psychology

A CONSIDERATION OF THE GENERALIZABILITY OF STUDENT RATINGS OF COLLEGE TEACHERS AND COURSES

A Dissertation Presented

by

HAROLD OLAF BETTENCOURT

Approved as to style and content by:

Alexander Pollatsek

Alexander W. Pollatsek, Chairman of Committee

Arnold D. Well, Member

Hari Haran Swaminathan, Member

William C. Wolf, Jr., Member

Jerome L. Meyers, Department Head
Department of Psychology

June, 1976

The law for the development of the self with respect to knowledge is this, that the increasing degree of knowledge corresponds with the degree of self-knowledge, that the more the self knows, the more it knows itself. If this does not occur, then the more knowledge increases, the more it becomes a kind of inhuman knowledge for the production of which man's self is squandered, pretty much as men where squandered for the building of the pyramids, or as men were squandered in the Russian hornbands to produce one note, neither more nor less.

--Kierkegaard

A C K N O W L E D G M E N T S

The author would like to express his appreciation to all those who contributed to this research project. I am especially grateful to the members of the dissertation committee, Drs. Pollatsek, Swaminathan, Well, and Wolf, for the time and effort they invested in reading the manuscript. Each of the members provided expert and resourceful comments that were important to the completion of the dissertation; particularly the committee chairman, Dr. Alexander Pollatsek. Considerable credit for the success of this research project is owed to Dr. Pollatsek for his patience and guidance during the conceptualization, operationalization, and ultimate conclusion of this dissertation and throughout my graduate career.

A very special thanks goes to Dr. Arnold Well, who, for reasons known only to himself, never opted to move his office to a more remote location of Tobin Hall to avoid the incessant questions of the graduate students next door. Dr. Well's open-mindedness was sincerely appreciated. There are, of course, many graduate colleagues of mine who contributed indirectly to the success of this project with their unfailing moral support and general raising of spirits, both psychologically and socially.

In closing, a very special extension of gratitude is given to Ms. Virginia Steel, my closest companion and confidant, for her patient proof-reading and typing of the original manuscript. And finally, I would like to express thoughts and feelings of goodwill, which I am sure Virginia shares, toward our devoted friend, Isaac F. Newton. His unfailing sense of bad timing often provided a source of laughter during the long hours of work required to write and type this dissertation.

H.O.B.
July, 1976

ABSTRACT

A Consideration of the Generalizability of
Student Ratings of College Teachers and Courses

(June 1976)

Harold O. Bettencourt, B.S., M.S., Ph.D.
University of Massachusetts, Amherst

Directed by: Professor Alexander W. Pollatsek

Undergraduates in economics, statistics, and German at the University of Massachusetts responded to pre- and post-course questionnaires pertaining to their demographic variables, pre- and post-course motivation-related items, and a post-course assessment of the quality of instruction/course. The pre-course questionnaire was administered during the first class period and the post-course questionnaire during the last week of classes prior to final exams.

Results suggest that student background items (i.e., "GPA", "college year", "major", "class size", and "optionality") are not related to student's teacher/course ratings. Pre-course motivation-related items (i.e., "amount expected to learn", "relevance" to intended career, desired "workload", "interest" in the course material, and "effort" willing to expend) were correlated with composite scores derived from specific ratings items. Students' assessments of these same items after the course was over (post-course motivation-related items) were strongly related to teacher/course ratings of "skill",

"rapport", "content", and "organization/structure" regardless of whether the ratings were general or specific in style. In addition, students' perceptions of the teacher's "enthusiasm" and ability to "entertain" were highly correlated with teacher/course ratings, particularly general rating items. Results observed within the economics classes were not found to be entirely consistent with results observed for statistic and German classes.

The implications of these findings are discussed with emphasis upon the use of students' teacher ratings in promotion and tenure decisions as well as teacher improvement.

An addendum to the primary research suggests possible sources of instability in factor structures derived from specific items within and between disciplines.

CONTENTS

LIST OF TABLES

The Generalizability of Students' Ratings of College Teachers

This dissertation is subdivided into multiple sections, each of which pertains to the generalizability of students' ratings of college teachers and courses. The sections are interdependent conceptually and must be read in the order that they are presented to obtain their full meaning. Since each preceding section serves to introduce the content and purpose of succeeding sections, only the initial sections require any introductory comments about their organization.

The first section, subtitled "A Selected Literature Review", is an objective presentation of the research literature pertinent to the generalizability problem and the current research. The objectiveness must be emphasized since the review is meant only to familiarize the reader with important issues and ideas necessary to understand the hypotheses developed in the second section, subtitled "Research Objectives". The "Research Objections" section requires that the reader be facile in the use of the concepts, results and criticism of the research as discussed in the literature review in order to understand the justification for the hypotheses and, therefore, the purpose of the current research. The hypotheses have been summarized at the end of the "Research Objectives" section. The reader may further benefit by pursuing these hypotheses

and then rereading the literature review.

A Selected Literature Review

Two important and related areas of educational and psychological research concerning student ratings of instruction are the generalizability and validity of these ratings. "Generalizability data describe the accuracy with which the data at hand represent data acceptable to a decision maker. For example, with what accuracy does a given set of ratings represent the ratings an instructor would receive were he to be evaluated by all students on all occasions of his teaching?" (Doyle and Whitely, 1974, p.260). Stated more simply, generalizability refers to the effect student background and motivation-related variables have upon students' ratings of teachers and courses. That is, how well can ratings be predicted given certain student characteristics. These student characteristics can vary as a function of the course (e.g., course "optionality") but should typically vary from student to student independent of the actual courses (e.g., student "interest" in the course material or perhaps, the "effort" the student expects to invest in the course). If student characteristics affect students' ratings independently of the quality of instruction, then these characteristics should be compensated for when

interpreting a faculty member's teacher/course ratings.

Validity, on the other hand, refers to the "meaningfulness" of students' ratings. Typically, researchers describe student rating validity in terms of internal consistency--how well the questionnaire items intercorrelated indicating some common underlying themes to the evaluation' instrument-- and in terms of strength of the association of ratings to some external criterion, such as, course grades, final exam scores, or exam scores "residualized" for initial ability (Costin, et al., 1971; Aleamoni, 1972 (a), 1974 (b)). Sullivan and Skanes (1974) point out, however, that validity might be better assessed by correlating student ratings with measures of course objectives other than the amount which students learn. For example, if one objective of a particular course (e.g., introductory psychology) is to stimulate student interest in the content area rather than insure that students acquire certain specialized knowledge, then an appropriate external validity criterion might be the assessment of students' willingness to elect another course or their choice to major within the discipline. The Sullivan and Skanes study brings to light a current and controversial issue concerning student rating of instruction: What is the appropriate validity measure? However, the content of this dis-

sertation does not focus upon the external validity issue (refer to Bettencourt, 1974, for an indepth review of the validity literature), but does deal indirectly with the internal consistency and/or the content validity issue. We will say that the instrument lacks validity to the extent that the items are clearly not measuring the actual quality of instruction, but are measuring other aspects of the course over which the teacher has no control: for example, if questionnaire items designed to measure an instructor's teaching skill correlate highly with students' reasons for taking a course (e.g., required or elected) or with students' pre-course attitudes about a course (e.g., perceived relevance or irrelevance of the material). In this instance, the instrument would lack more than content validity, it would also lack generalizability. That is, the questionnaire items would be measuring student facets unrelated to teacher performance and, therefore, the ratings would not be generalizable over student characteristics.

Evidence for and against generalizability. Even though the actual concept of "generalizability" has been infrequently mentioned in research literature, many researchers (Costin, et al., 1971; Aleamoni, 1972 (a), 1974 (b); Granzin and Painter, 1973; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; Sullivan and Skanes, 1974; Doyle

and Whitely, 1974) have examined or described the relationship between student background characteristics and the predictability of ratings of teaching ability. The question is, to what extent do particular background facets of a rating population affect these students' opinions of college teaching? Typically, the degree of predictability is couched in terms of the proportion of variance in a general teacher rating item accounted for by several background items (square multiple correlation). For example, Doyle and Whitely (1974) show that as much as thirty-six percent of the variance of rated "course effectiveness" can be predicted by a linear combination of background and motivation-related variables, such as "required course", "year in school", "liking for subject", to mention a few.

The fact that research (see Costin, et al., 1971; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; Aleamoni, 1974) has repeatedly demonstrated that most background variables typically do not correlate significantly with teacher ratings is usually cited as support for the generalizability of students' ratings of their college teacher. (These background variables are discussed below.) However, the fact that these variables are commonly unrelated to students' teacher ratings renders them less important to the current research, and therefore, the results from current empirical literature pertaining to

these variables will be reviewed in a cursory manner.

Sex of the rater usually has little or no affect upon teacher ratings for either general or specific items (Costin, et al., 1971; Aleamoni, 1974 (b); Granzin and Painter, 1973; Centra, 1973 (a,b); Perry and Bauman, 1973; Aleamoni, 1972; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; Kulik, 1974). However, Doyle and Whitely (1974) report that student sex correlates significantly ($r \leq .18$) with some general items (e.g., "liking for person", "how much learned", and "how motivating"), although the variance accounted for is less that three percent. Elmore and LaPointe (1974) suggest that females may rate instructors slightly higher than males on "structure" variables.

Year in college, on the other hand, has been found to have a slight positive relationship with teacher ratings (Costin, et al., 1971; Doyle and Whitely, 1974; Kulik and Kulik, 1974; Kulik, 1974 ; Kulik and McKeachie, 1973; Kohlan, 1973; Aleamoni and Yimer, 1974; Aleamoni and Graham, 1974; Perry and Bauman, 1973; Aleamoni, 1974 (b)). That is, as students progress from freshman to senior to graduate college years, ratings of instruction tend to become more positive although they still have low predictive power ($r \leq .15$). Grasha (1972) found no relationship between "year in college" and student ratings at the University of Cinncinati.

Students typically rate courses within their major area of study slightly higher than non-major courses (Costin, et al., 1971; Grasha, 1972; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; Aleamoni, 1974 (a, b)). However, Granzin and Painter (1973) report no increase in predictive power from knowing students' majors.

Increased class size typically either lowers teacher ratings (Costing, et al., 1971: Perry and Bauman, 1973; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; and Kulik, 1974) or has no apparent effect (Aleamoni, 1974 (a, b); Aleamoni and Graham, 1974; Grasha, 1972).

Grade point average (GPA) tends to correlate near zero with students' ratings (Aleamoni, 1974 (b); Grazin and Painter, 1973; Kulik and Kulik, 1974), while course grades, even though closely related to GPA, have been reported in early literature (Costin, et al., 1971) to correlate positively with rating outcomes ($r \leq .30$). However, the course grade item is more closely related to validity issues and will be discussed later as being highly controversial in that context.

Because each of the background items just reported generally accounts for less than nine percent of the variance in any teacher rating item, it seems reasonable to argue that evaluations can be administered without faculty members having to fear that these student back-

ground variables will contaminate results. Teaching ratings are generalizable across these variables.

There are, however, several other student variables which significantly correlate with student ratings of teaching. Optionality (whether the course is required or elected) and motivation-related variables have repeatedly correlated with students' ratings of college teachers/courses.

Research consistently shows "optionality" correlating significantly ($r \leq -.32$) with students' opinions of teachers and courses. In effect, as the percentage of students taking a course as a requirement increases, the rated quality of the overall course decreases (Costin, et al., 1971; Doyle and Whitely, 1974; Granzin and Painter, 1973; Grasha, 1972; Kulik and Kulik, 1974; Kulik and McKeachie, 1974). Gillmore and Brandenberg (1974) have provided the most thorough demonstration of the reduction in ratings resulting from the "optionality" variable although their research procedures were not methodologically perfect. These authors collected student rating data on 763 class sections at the University of Illinois at Urbana-Champaigne. The rating instrument was composed of specific items, eight per factor, for five factors: "general course attitude", "method of instruction", "course content", "interest/

attention", and "instructor". The independent variable in the ANOVA was the percentage of "required" students per class section. For example, a typical group would be composed of all courses having between forty and fifty-nine percent "required" students, another would be those having twenty to thirty-nine percent, and so on. Results clearly indicate a significant decrease in mean ratings between each of the five experimental groups as the percentage of "required" students increases from group to group. The change in average of rated instructor/course qualities ranged between eight percent and nineteen percent on a four-point scale for ratings of "general course attitude" and "instructor". It would seem that "optionality" clearly affects the generalizability of students' ratings of college instruction: the greater the percentage of "required" students taking a particular course, the poorer the average ratings of instructor/course quality will be, somewhat independent of teacher performance. However, Gillmore and Brandenburg did not report one piece of evidence essential to this conclusion: the effect of course "optionality" within each course section. As a matter of fact, this author did not find a single recent study reporting such data. Furthermore, Costin, et al., (1971) reviewed studies as far back as the 1930's and did not report such data either.

Therefore, it appears as if "optionality" correlates highly with teacher ratings, but additional within-class evidence would seem needed. In any event, the question of course "optionality" will be investigated within and between class sections in this research.

Motivation-related questionnaire items. Motivation-related variables, unlike background variables, relate to students' attitudes about a course which might change during the semester as a result of actually taking the course. For example, a student's "interest" in the course, perceived "relevance" of the material, perception of "amount learned", and "effort" expended or willing to expend during the course might predict to a great extent the rating pattern a particular instructor and course receive at the end of a semester. This suggests that the correlations between any or all of these motivation-related items and teacher ratings would be quite interesting when considering the generalizability issue. Even more important than simply demonstrating that the relationship exists would be determining the extent of predictability of rating outcomes by knowing students' motivational attitudes. In other words, what kinds of attitudes should a teacher try to instill in his students to insure high teacher ratings? Only recently have researchers in education and psychology begun to investigate the contribution of motivation-related items to the problem of generalizability of teacher ratings.

Granzin and Painter (1973) at the University of Utah had 637 students in seventeen different courses and eleven departments rate college coursesboth early and late in the semester. In addition to the usual background variables (e.g., "class size", "college year", etc.), several "static" and "dynamic" variables were measured. The static variables which were rated only once at semester's end were "effort expended", "course importance", "contribution to general education", "contribution to vocation", "interesting/entertaining course", "ease of course", and "grade expected". That is, students' pre-course attitudes were assessed and then subtracted from their post-course attitudes. Therefore, dynamic ratings are simply raw difference scores. (The use of raw difference scores is generally considered to be methodologically unsound by many researchers, and it is a point deserving serious criticism in the Granzin and Painter study. Refer to Crombach and Furby [1970] and Cohen and Cohen [1975].) Each of these motivational variables (both static and dynamic) were correlated with the three separate general course ratings of quality of the "overall course", overall "course content", and overall "instructor" ability.

Results revealed significant positive correlations for both static and dynamic variables-- "contribution to general education", "contribution to vocation", "inter-

esting/entertaining", "effort", "importance"--with the three general rating items (r's ranged from .39 to .62, $p<.001$). This means that ratings of college teachers and courses, as reported by Granzin and Painter, are heavily influenced by motivation-related variables: they are not generalizable over these items. In general, however, static variables produced larger correlations than did dynamic variables. Of particular interest is the fact that all significant correlations reported are substantially larger than those obtained by correlating background variables with general rating items.

A second important issue addressed in the Granzin and Painter article is the extent to which each general rating item can be predicted by the best linear combination of motivational variables (multiple regression analysis), and which of these motivational variables will be most important (as determined by the regression coefficients).

A stepwise multiple regression analysis was performed using each general item as a criterion variable to be predicted by the motivation-related items. The most important motivation-related items for predicting "overall course", "course content", and "instructor" ratings were "interesting/entertaining", "contribution to general education", and "contribution to vocation". The total vari-

ance accounted for by the best fitting linear equation for the three criterion variables were 57.3% ("overall course"), 42.9% ("course content"), and 44.9% ("instructor"). In each of the regression equations, "interest/entertainment" loaded most heavily (B=.32), nearly twice as much as any other predictor variable. This result coincides with the hypothesis offered by Naftulin, Ware, and Donnelly (1974), and Ware (1974) who suggest that a teacher who is very entertaining may "seduce" students into providing higher ratings even though the teacher has relayed little substantive material. This, if indeed true, would be cause to seriously reconsider the utility of student ratings since the evaluation instruments would lack generalizability and validity.

The major contribution provided by the Granzin and Painter study is its novel approach to the generalizability problem, that is, the introduction of static and dynamic motivation-related variables. However, the study has serious methodological problems deserving criticism. First, both background and motivation-related variables were only correlated with general teacher/course rating items. Quite different results might be expected if the teacher/course rating items were specific items (e.g., "the teacher spoke clearly") or were factor clusters of specific items rather than general items. Secondly, "interest/entertain-

ment" should appear as two separate items on rating instruments. The rationale for this statement is the following, as Naftulin, et al., suggest, teachers who are very entertaining may receive high ratings even though little substantive material has been made available. Under these circumstances, rated interest in course material probably would be low. However, these issues cannot be resolved on the basis of either Granzin and Painter's data or Naftulin's, et al., data. Additional research is needed before any substantial conclusions can be reached.

A third problem area concerns Granzin and Painter's data analysis. They collapsed their data across several courses, teachers, and disciplines, and therefore the relationships they report between variables may not be true or consistently found within course sections or single discipline groupings. That is, they have confounded teacher and student differences in their data analysis. In addition, they used raw difference scores in their analyses which, as explained in a previous paragraph, is reason enough to seriously question their results and conclusions. In the current research, portions of Granzin and Painter's study will be replicated taking care to correct their methodological flaws, particularly the use of pooled data and raw difference scores. However, the Granzin and Painter study is only one of several studies which helped

to provide the impetus for this research.

Doyle and Whitely (1974)[1] have reported a study similar to that of Granzin and Painter's using background and motivation-related variables as correlates of both general and specific rating items (factor groupings). These researchers asked 174 students in twelve undergraduate French courses at the University of Minnesota to rate the quality of their graduate student instructors on seven general items (i.e., "liking for person", "general teaching ability", "attitudes about teaching", "how much learned", "how motivating", "overall teacher effect", "overall course effect"), forty-nine specific items (e.g., "made it clear how each topic fit into the course"), four background variables (i.e., "year in school", "sex", "MSAT ability", "required course"), and four motivation-related variables ("liking for subject", "how hard tried", "conscientiousness" in filling out questionnaires, "usefulness"). All course sections had common exams, texts, and syllabi to insure that students were exposed to the same materials, thereby reducing the possible error variance which might arise if these variables were not controlled. (This was also a major problem in the Granzin and Painter study.) The data were subjected

---

[1] This study was published after I had designed my study and had collected the pilot data. Therefore, portions of my research will serve to replicate Doyle and Whitely's results with several extensions (e.g., additional motivation-related variables).

to separate "across-sections" and "between-sections" analyses. Across-sections analysis refers to those on the item-level for all 174 subjects as a combined group. For between-sections analysis means were computed for each item within each class of the twelve course sections. These means were then used for correlational analysis. The small number of students (approximately fifteen) in each class section prohibited consideration of teacher rating analysis within-class sections (a serious methodological problem in this study and that of Granzin and Painter).

The results of the across-sections analysis exhibited the following relationships. Some of the background variables were significantly correlated with one or more of the general rating items (seven significant correlations out of the twenty-eight possible). However, the largest of these correlations accounted for only four percent of the variance of the general items ("MSAT ability" with "overall course effectiveness"). These moderate or near zero correlations between background variables and general ratings are consistent with the research literature reviewed earlier in this proposal, particulary the Granzin and Painter results. Thus, students' teacher ratings appear to be generalizable over these background variables.

All but one motivation-related variable ("liking for subject") correlated positively and significantly ($P<.01$) with all the general teacher/course rating items (the variance account for, r^2, ranged from .05 to .49). "Liking for the subject" exhibited no relationship with the general item, "liking for the person", which indicates that students' positive feelings for their courses were not the result of teacher popularity. The relationship of "liking for the subject" was substantial ($.04 \leq r^2 \leq .25$) with the remaining general rating items (e.g., "general teaching ability", "how much learned", etc.) indicating the importance of this motivation-related variable in students' assessments of their college teachers. These data do not argue against the generalizability of the teacher ratings since the study used post-course motivation-related variables over which the teacher should have control. Therefore, the student attitudes would be expected to reflect their feelings about the teacher.

The between-sections data, while exhibiting two nearly significant correlations ("year in school" with "general teaching ability", $r=.40$, and "MSAT ability" with "attitudes about teaching, $r=-.39$; $p<.05$ for $r \geq .49$) must be viewed with caution since these correlations are based upon a sample size of twelve (sections). For this reason, further discussion of the between-sections data will not

be considered.

Additional analyses of the extent of the relationship of the background and motivation-related variables (predictors) with the seven overall ratings were achieved using canonical correlational analysis and multiple regression analysis. A "teaching quality composite", made up of the linear combination of "general teaching ability", "attitudes about teaching", and "general teaching ability" was the criterion variable in the canonical analysis. The results of these analyses further support previously reported findings: background variables are not related to general teacher ratings, whereas motivation-related variables are substantially related.

The canonical correlation accounted for twenty-six percent of the variance in the "teaching quality composite", indicating that the motivation-related variables "how hard tried" (B=.51) and "usefulness" (B=-.53) were the most important predictors. In addition, the regression analysis accounted for 18%, 50%, 38%, and 36% of the variance in the remaining four criterion variables, "liking for person", "how much learned", "how motivating", and "course effectiveness", respectively. The most important predictor variables resulting from multiple regression on the four remaining general rating items were the following: "conscientiousness" in filling out questionnaires loaded significantly upon

"liking for person" (B=.30); "how hard tried" loaded on both "how much learned" (B=.50) and "how motivating" (B=.32); "liking for subject" also loaded heavily on two items, "how motivating" (B=.21) and "course effectiveness" (B=.50). None of the other beta weights were significant (usually near zero) indicating a lack of usefulness of the remaning variables in predicting students' general ratings of teachers and courses.

Whereas the results reported above demonstrated a consistent and strong relationship between motivation-related variables and general teacher rating items, the relationship of these same motivation-related variables with specific items was different. In fact, the motivation-related variables were not related to any of the five factors derived from the fourty-nine specific items. For example, the factors, "attitude toward students" (rapport), "expositional skills", "motivation of interest", "stimulation of thinking", "generalization of content", were checked for generalizability (independence from background and motivation-related variables) using Dwyer's factor extension method. Dwyer's method "permits the estimation of factor loadingsof new variables on the factor structure of prior set with the advantage that the extended loadings do not alter the initial structure." (Doyle and Whitely, 1974, p. 266). The only item loading significantly on

the exi sting factor structure was rated "liking for subject" which loaded .41 (loadings $\geq$ .325 implied $p < .05$) on the "motivation of interest" factor. These results suggest that factors derived from specific items are relatively independent of students' background and motivation-related variables. That is, the factors are independent of these student facets and, therefore, are more generalizable. On the other hand, the ratings of general teacher/course items could be predicted to a large extent by the motivation-related variables and some background variables, implying a lack of generalizability for general rating items. These results, indicating different levels of generalizability for general and specific items, suggest that evaluation instruments might be more useful if composed of a series of specific items which tends to confirm similar specultation by other researchers (Grasha, 1972; Kulik and Kulik, 1974; Kulik and McKeachie, 1973).

As important as Doyle and Whitely's results appear to be to evaluation instrument development, their data and findings must be considered with caution because of some serious methodological problems. First, the primary mode of lecture presentation (and of student learning) was by video tape on closed circuit television twice per week, supplemented by one discussion section per week with one of twelve graduate teaching assistants. Because

of the method of presenting lectures and the ancillary role of the graduate assistants, the Doyle and Whitely results probably should not be generalized to any university faculty population. The students were not actually rating the primary source of learning-- the video-taped lectures--but rather their discussion section leaders. Furthermore, it might be that this kind of lecture presentation dramatically affects (either positively or negatively) students' motivation to learn or their interest in the course. If this were true, perhaps we could expect motivation variables to be stronger predictors of both general and specific teacher rating items.

In addition, one would expect large group differences in the teaching ability of faculty members and of graduate students along several dimensions (e.g., their experience as teachers, or their knowledge of the subject matter) as well as differences in the students' perceptions of the two kinds of teachers as educators and persons. (This might affect "rapport" items and consequently influence many other of Doyle and Whitely's results.) In effect, students probably rated these graduate student teachers on their ability to complement the video-taped lecture performance of the faculty member. Rating patterns and relationships would probably appear very different if students were asked to rate the primary lecturer rather

than discussion leaders.

Another problem area is the small sample size. Because of the small sample size, Doyle and Whitely could not effectively analyze the relationships within sections to assess rating pattern consistencies between sections and compare these results to the overall data analysis picture. Each of these criticisms is important in and of itself, however, when considering them collectively, they suggest a need to replicate Doyle and Whitely's basic experimental study using university faculty, not graduate students or video-taped lectures. Furthermore, the process of replication should include some design changes, particularly the selection of rating items used to assess motivation-related variables. While those items reported by Doyle and Whitely are important, others used by Granzin and Painter also seem appropriate and interesting (e.g., the "entertainment" issue).

Many issues have been presented in this "selected literature review" which the reader should carefully note. Several of the more important will be restated for the reader's benefit. First, background variables seem unrelated to students' general teacher/course ratings, whereas motivation-related variables are strongly related. Second, some evidence was reported suggesting that the relationship of background and motivation-related variables with

specific teacher/course ratings may be different. This might provide some insight into which type of rating instrument (composed of general or specific items) is more desirable due to its generalizability. Third, it seems probable that general teacher/course ratings are related to students' pre-course motivation-related variables. Such a finding would be cause for a serious reconsideration of the student-teacher rating process.

Finally, the state of the literature, as reported here, clearly suggests the need for additional research because of methodological flaws in each research study reported. For example, one such problem was the failure of any research study to do within-class data analyses. In that researchers have been interested in the generalizability of students' teacher ratings over various student facets, the within-class analyses are essential to answering questions concerning the generalizability problem. The present research corrects this methodological problem and others.

SECTION I
RESEARCH OBJECTIVES

The current research will replicate and extend portions of the Granzin and Painter (1973) and Doyle and Whitely (1974) studies, taking precautions to correct the major methodological flaws in each. (The major design problems, as discussed in the preceding literature review, will be restated briefly with the introduction and explanation of each type of experimental analysis.) Basically three distinct areas of interest will be investigated in this research. The first is the effects of student background and motivation-related variables upon both general and specific (linear combinations of specific items) teacher/course rating items within a single discipline (economics). As was pointed out earlier, McKeachie, Lin, and Mann (1971) suggest that instruments composed of specific ratings might be less predictable from these student variables than instruments composed of general items.

The second area is an assessment of the consistency of the major findings from the economics data across other disciplines (statistics and German). The consistency of results across disciplines is important in determining whether or not the subject matter of a particular course or discipline might be related to certain distinct tendencies in teacher/course rating outcomes.

The third area is an assessment of the consistency of the factor structure derived from specific rating items within a discipline (economics) and between disciplines (economics versus statistics versus German). This will be determined using maximum-likelihood factor analytic techniques (Joreskog, 1966, 1967, 1971.)

The results of the first and second areas of research will be discussed in light of the usefulness of rating data for faculty members and administrators, particularly when rating data are used in personnel decisions (e.g., promotion and tenure). The results of the third area of research will be discussed separately in the addendum to the main body of the dissertation. The focus of this discussion will pertain to the utility of rating instruments composed of general or specific items.

The relationship of background and motivation-related student variables with general and specific ratings of teacher/course effectiveness. A lack of generalizability can be thought of here in terms of the "differences in evaluations that might arise from sampling matters: differences due to the occasions of teaching that are evaluated, differences due to the circumstances surrounding the evaluation and so forth." (Doyle, 1975, p.71). Therefore, one question to be considered is how much student population characteristics (which are not controllable by

the teacher) affect teacher ratings. Most of the literature reported that deals with this issue has been concerned with the effect of student background variables (e.g., Costin, Greenough, and Menges, 1971; Kulik and McKeachie, 1973; Kulik and Kulik, 1974; Aleamoni, 1974 (b); Doyle, 1975). Correlations between these variables and general teacher/course ratings are typically small ($|r| \leq .15$) or show no relationship at all. The exceptions to this rule are ability measures, like final exam grades or course grades which usually produce slightly largercorrelations. The more popular (i.e., most frequently cited in research literature) student background variables are student "sex", "college year", "major", "class size", "optionality", "GPA", "final exam grade", and "course grade". However, the studies above produced inconsistent patterns of relationships between these latter background variables and general teacher/course ratings. Probably the most parsimonious description of these results can be provided through a description of variance accounted for by the relationships between the different student variables and teacher/course ratings. None of the variables reported here account for more than nine percent of the variance in any general rating of teacher/course effectiveness (e.g., "how would you rate your teacher's overall ability?"), and most often they explain no significant variance whatsoever.

The exceptions to this rule of thumb are course "optionality", which consistently accounts for slightly more variance, about ten percent, and "final exam scores", which typically accounts for nine percent. One can conclude then, that for the most part, general teacher/course ratings are generalizable over student background and/or ability measures. (The reader should note that generalizability as used here implies a continuum from complete generalizability, or no correlation between ratings and student variables, to a complete lack of generalizability, a perfect correlation between ratings and student variables. In this research, the more generalizable the rating items the more useful the rating instrument.)

Recently, however, some researchers (Granzin and Painter, 1973; Doyle and Whitely, 1974) have introduced a different student typology, that is, student motivation-related variables, as predictors of general teacher/course ratings. (It should be noted that course "optionality" has been considered to be a motivation-related variable by Doyle (1975) rather than as a background variable as most researchers have chosen to label it. It would seem to be both.) Granzin and Painter assessed variables like students' perceptions of the interest/entertainment value of course lectures and the contribution of course material to their general education while Doyle and Whitely assessed

variables such as students' perceptions of the amount of effort they expended learning course subject matter and their liking for the subject area, as predictors of general teacher/course ratings. The general ratings that are most important here and common to both research papers were "overall assessment of the course content" and "teacher effectiveness". The resulting correlations of these motivation-related variables with the general teacher/course ratings were, typically, much larger in magnitude ($.24 \leq r \leq .63$) than those already reported for background variables.

Unfortunately, even though the correlations were large, the results were not very meaningful since the parameter estimates are based upon pooled data matrices. Pooling the data matrices confounded the effects of the teacher and the course material with the student ratings. Interpreting results from pooled data analyses can be misleading. One cannot be sure whether or not the students' ratings of an instructor or course were due to teacher performance, course subject matter or both. Ideally, when one wants to examine the generalizability of students' teacher ratings, the course material should be identical from class to class, particularly if consistency of rating patterns (replicability) is important. The less control a researcher has over the course content, course syllabus

text material, exams, and so on, the more difficult it is to determine the causes of rating scale contamination or exactly what the rating items are measuring (e.g., actual teacher performance or some halo effect). Consequently, the best data a researcher probably can hope to obtain in a field study (like the study presented in this dissertation) would occur when several different instructors would teach large sections of the same course using common syllabi, exams, etc. which would allow the researcher to "look" within (and between) class sections for significant and consistent relationships between student variables and rating outcomes. Since both Granzin and Painter and Doyle and Whitely had sample sizes that were too small ($Ns<17$) to perform a meaningful within-groups analysis, they where forced to pool over teachers which led to the confounding discussed above.

A major advance of this research over previous efforts is the use of course sections (in economics) containing large numbers of student ($Ns\leq 195$), allowing for more stable estimates of the zero-order (Pearson product moment) correlations of background and pre-course motivation-related student variables with both general and specific teacher ratings. An additional improvement is that teachers in this study are all full-time university faculty members, rather than graduate student teachers leading discussion sections

of an ancillary nature. Furthermore, this study represents the first instance where general ratings and factors composed of specific items have had corresponding meanings permitting rating comparisons of the relationships of student variables with each type of teacher/course rating item (general versus specific). The results of these correlational analyses will be essential in determing the usefulness of certain student variables as predictor variables in later multiple regression analyses.

As stated earlier, Granzin and Painter (1973) were first to introduce the notion of student motivation-related variables (e.g., "interest in course material", "perceived importance of the course", "effort expended in the course", etc.) and to suggest that these variables were significantly correlated with general teacher/course ratings. Their data suggest that these student variables can be dramatically related to general teacher/course ratings (e.g., the rated "overall course quality"), as indicated by their reported correlations which were much larger in magnitude than those typically reported for background variables ($.25 \leq r \leq .63$). However, there are two major areas of their research which should be replicated, taking care to correct several methodological problems. For example, some effort should be made to ascribe meaning to motivation-related variables, and "difference score" should be eliminated as an independent

variable. Each of these problem areas will be considered in turn.

Granzin and Painter introduce student motivation-related variables without any justification as to why these variables might be interesting or what advantages there might be to adding a new set of variables to those already being used by researchers (e.g., background variables). Therefore, another objective of the current research was to provide some meaning, if possible, to motivation-related variables in addition to determining the relationship between motivation-related variables and teacher/course ratings.

One might hypothesize that the meaning of student motivation-related variables could be couched in terms of background variables. For example, it could be that students' "interest" in a course is explainable by some linear combination of background variables such as "GPA", "optionality", "college year", "major", etc. If so, then motivation-related variables have the advantage of presenting a unified conceptualization of the relationship between certain combinations of background variables and teacher/course ratings. This hypothesis would seem more likely to be true if students' motivation-related attitudes were assessed during the first class period, before the students actually had any course experience. In other words, the attitudes students' bring into the classroom are, initially, independent of the teacher's

classroom performance and probably more explainable by aspects of the course which do not change during the semester (e.g., "optionality"). If these latter statements are true, then little information would be gained by investigating motivation-related variables, since it has frequently been reported that background variables do not correlated with teacher/course ratings.

Alternatively, it can be hypothesized that student motivation-related variables are relatively independent of student background variables, and probably represent important criteria used by students to assess a teacher's class room performance. This has been indirectly suggested by much of the literature already reviewed. In any case, one can test these hypotheses in the present data. If the data support the latter hypothesis (as expected from the literature) then the uses of students' teacher rating data in personnel and other decisions would have to be carefully re-evaluated. The implication is that student rating data may have limited validity.

To summarize briefly, the objective here is to determine if students' pre-course motivation-related variables can be explained as linear combinations of background variables. The principal method of analysis here will be a stepwise regression analysis within each economics class. The background variables used as predictors will be "GPA", "major",

"college year", "class size", and "optionality." Linear combinations of these variables will be used to predict (explain) the following student pre-course motivation-related variables: the "amount expected to learn"; the "relevance" of the course content to students' planned careers; the desirability of the "workload"; student "interest" in the course; and the amount of "effort" students expect to invest in learning course material.

If the data should indicate, unexpectedly, that student motivation-related variables are explainable as combinations of background variables, then it will be of some interest to determine which of the motivation-related variables are more strongly related (if at all) to teacher/course ratings. However, if the background and motivation-related variables are not significantly related as anticipated, then several additional analyses will attempt to determine precisely the extent of the relationship between the motivation-related variables and students' teacher/course ratings.

To be more specific, a group of analyses will determine which type of teacher/course rating items, general or specific, are more empirically generalizable over student background and motivation-related variables. For the most part, this is a question of "degree of generalizablity"; the type of rating item least contaminated by student

variables will provide the more useful data for personnel decisions (Centra, 1973; Aleamoni, 1974 (b); Doyle, 1975). As mentioned earlier, Doyle and Whitely (1974) reported results directly related to this question. However, it was also pointed out that their analysis used a pooled data matrix which confounded the effects of course instructor and course material. Furthermore, the factor composites derived from their specific items failed to coincide in meaning with their general rating items. Therefore, a comparison of the generalizability of general and specific (factor composites) items having similar meanings could not be made. This certainly added to the difficulty in interpreting their empirical findings, and especially their conclusions about the relationship of background and motivation-related variables with the two types of rating items.

In the present study these problems have been eliminated. For example, three of the general rating items used (students' assessment of overall teacher ability or "skill", course "organization/structure", and teacher-student "rapport") were selected because they were parallel to three of the five dimensions commonly derived from specific teacher/course items. The selection of the specific items themselves was based upon their occurrence in current research literature as loading on one of five factors; "skill", "organization/

structure", "rapport", "workload", and "general course attitude" (refer to Appendix C for a summary of the applicable studies). For example, specific items, such as "the instructor was well-prepared for each lecture" and "the instructor expressed ideas clearly", typically load on the "skill" dimension for an orthogonal factor solution under varimax rotational procedures. It is, consequently, a simple matter to compare the functions resulting from a regression analysis using the same student characteristics to predict both the general rating of teacher ability and the composite "skill" dimension. Such comparison could be made on the basis of variance accounted for in each of the criterion variables (e.g., "skill", "rapport", etc.) by the linear combination of predictor variables (i.e., background and motivation-related).

To summarize briefly, it is expected that specific ratings, and therefore, factor composites of these ratings, will be less contaminated (predictable) by student variables than general rating items of similar meaning. For example, students might possibly emphasize some inappropriate criteria, such as the "entertainment" value or the "optionality" of a course, when assessing overall teacher ability. This is probably less likely to occur if the student is asked to rate a specific behavior, e.g., "the

instructor expressed ideas clearly". This will also be tested in the current data.

As indicated earlier, several points should be made concerning the use of "difference" scores in the Granzin and Painter (1973) article, which lead to the last two analyses of this research. The question of how much motivation-related variables might change as a result of taking a particular course and the extent to which these variables will predict teacher ratings is an important one. However, the use of raw "difference" scores to measure these effects is empirically undesirable. Several authors (Cohen and Cohen, 1975; Crombach and Furby, 1970) have cautioned against the use of raw "difference" scores as a measure of change. The danger in using these scores stems from their unreliability: in general they cannot be expected to correlate substantially with anything else (Cohen and Cohen, 1975, p. 64) since they consist mostly of measurement error. The unrealiability of "difference" scores and the problem of removing the contaminating effects of the pre-score from the post-score can be easily corrected. For example, the correlation between a "difference" score and another variable can be achieved by simply partialling out the pre-score from the post-score, and then correlating this residual with the other variable (Cohen and Cohen, 1975, p. 381). This partialling procedure was achieved by using traditional multiple regression techniques.

A major criticism of the Granzin and Painter (1973) study resulted from these researchers correlating raw "difference" scores with general teacher/course ratings. In effect, the "difference" scores correlated less substantially with the general ratings than did the post-course motivation-related scores alone. Based upon these results, the authors concluded that positive changes in attitudes were transferred by students into positive teacher/course ratings. However, the evidence reported above concerning "difference" scores suggests that the Granzin and Painter results might be spurious, a result of methodological problems inherent in the "difference" score approach. In the present study, these methodological problems have been corrected. The relationship of student attitude change to teacher/course ratings will be re-examined.

To determine the strength of changes in motivation-related variables as predictors of end-of-semester teacher course ratings the following analysis will be performed. The effects of both background and pre-course motivation-related variables will be partialled from the end-of-semester general and specific ratings in a hierarchical manner. That is, the background and pre-course motivation-related variables will be partialled from the regression equation prior to entering the post-course variables of identical meaning. This procedure will be followed for both general

and specific items. In effect, these analyses will correlate linear combinations of students' residualized (change in attitude) post-course motivation-related variables with teacher/course ratings. These changes in attitude are, presumably, due to the impact of the teacher and course material upon students enrolled in a course. Therefore, one would expect from the literature and from intuition that residualized motivation-related variables would be strongly related to both general and specific rating items, but much more so to general rating items. The stronger relationships expected for general items is suggested by Doyle and Whitely's data reported earlier as well as speculation by McKeachie, Lin, and Mann (1971).

The last analysis is intended to determine how important post-course motivation-related variables are as predictors of general teacher/course ratings. In addition, two other variables will be used in the equation to determine their ability to augment this prediction of ratings. The two additional variables are the student's perceptions of a teacher's "enthusiasm" and the "entertainment" value of lectures. The idea, then, is to determine how much variance of the general ratings can be predicted after the effects of specific ratings have been partialled from each general rating item.

In the present data, linear combinations of specific

rating items were used as the first set of predictors of general teaching "skill", teacher-student "rapport", "content" coverage, and course "workload". If students rate a teacher/course on the basis of performance alone, then the specific items should account for nearly all the variance of the general ratings. Alternatively, if students base their general ratings upon other teacher behaviors, such as "entertainment" level and instructor "enthusiasm", then the post-course motivation-related variables and other variables should account for a substantial portion of the variance in the general ratings. This would be evidence to suggest at least a cautious use of general teacher/course rating items. Perhaps, depending upon the strength of the relationships, an even stronger recommendation could be made: a discontinued use of general rating items altogether.

Before providing the reader with a recapitualtion of the hypotheses as discussed above, a cautionary note is in order. McKeachie, Lin and Mann (1971) have argued that, in "field studies" such as this one, equally or more important than observing a single large, significant result is that the results obtained in one student population (class section) replicate within other student populations (particularly classes in other disciplines). The reason that replication of results is so important to applied problems can be described in the following man-

ner. To observe a particular result in only one student population is insufficient evidence to conclude that the observed result is generally inherent in all student populations. This is particularly true when doing evaluation research, since there are always a large number of uncontrolled variables which could influence results from population (classroom) to population (classroom). However, if a result is found to be significant within various rating populations, then a researcher can be reasonably sure that the results observed are meaningful and probably common to all student populations.

The present research was designed to allow an investigation of the replicability of results within several sections of economics that had similar, if not identical content coverage, syllabi, text material, and some commonality of exams. Furthermore, the present data will also allow a consideration of the replicability of the results reported in the economics data within other disciplines (e.g., statistics and German). Therefore, all of the analyses performed within each economics class section will also be performed in statistics and German classes, sample sizes permitting. (Unfortunately, the natural selection process, or the amount of incomplete data obtained in statistics and German classes, as a result of subjects being absent on one or both days when the questionnaires were administered, produced sample sizes too small

to allow within-class analyses in these disciplines). The results of these analyses are reported in Section II of this dissertation under the heading of "Grouped Data Analyses".

For the reader's convenience, an abbreviated summary of the hypotheses and analyses describing the purposes and objectives of this dissertation research are now presented. The topic areas and hypotheses will be restated in the order in which they were orginally introduced.

1. The first analysis focuses upon the question of generalizability of both general and specific teacher/course rating items. To be more specific, the predictive power of two different types of pre-course items relating to certain facets of the student rating population will be examined; a) background items, and b) motivation-related items. A major hypotheses suggested by the literature and by intuition is that specific rating items should be more generalizable over both background and motivation-related items than general rating items. This will be examined using Pearson correlational and regression analytic techniques.

2. The second analysis focuses on the predictive power of the motivation-related items and especially whether they predict any of the variance in the general course ratings beyond that accounted for by student background items. This will be ascertained using a hierarchical regression analysis with items representative of student background character-

istics entered into the equation first. However, as suggested from the literature reviewed, neither type of pre-course item (e.g., background or motivation-related) is expected to significantly predict specific items.

3. The third analysis focuses upon the degree of control an instructor has upon student motivation and the predictive power such motivational changes might have upon teacher/course ratings, both general and specific. It would seem from the literature reviewed that the residualized (i.e., with the effect of pre-course items partialled out) post-course motivation-related items should predict considerable variance in the general teacher/course ratings beyond that predicted by background items. Alternatively, specific rating items should be less predictable as criterion variables in similar analyses.

4. The fourth data analysis focuses on a different aspect of the validity of teacher evaluations. If teacher "entertainment" and "enthusiams" remain significant predictors of teacher ratings (whether general or specific) after the effects of performance-related variables (e.g., "skill", "organization/structure", "content", and "rapport") and motivation-related items have been partialled out, then one might view these ratings with some skepticism.

The literature suggests that general rating items will be more predictable by teacher "entertainment" and "enthusiasm" than groups of specific items corresponsing in meaning.

Method

Subjects and Administration. Courses in four sections of economics, six sections of statistics and five sections of German were rated on common pre- and post-course questionnaires at the University of Massachusetts. The pre-course questionnaire was administered to students at the beginning of the second class meeting. Administration of the post-course questionnaire took place during the last regular class, before the beginning of the final exam week. Both questionnaires could be completed in less than ten minutes.

Total sample sizes by subject area were 962, 204 and 100 for economics, statistics, and German, respectively. The student population in each of the courses was composed of freshmen, sophomores, juniors, and seniors taking these courses as a requirement or as an elective. (Refer to Appendix B.) The teachers were all full-time University of Massachusetts faculty members, meeting three class hours per week with their students. In the economics and German courses teachers gave similar or common final exams. All sections of German had, in addition, common texts and syllabi. Portions of the economics syllabi were also common (covering the same substantive topic areas), using identical texts and several common test questions on the final exams. On the other hand, the statistics sections used common texts, but did not follow common syllabi, nor did the faculty members collaborate on a final

exam.

Instruments. The pre-course questionnaire asked students, in addition to background characteristics, what they anticipated gaining from the course in terms of these six categories: a) "grade satisfaction", i.e., what grade the student would be satisfied with; b) "amount expected to learn"; c) "relevance" of course material to planned occupation; d) "workload" desired; e) "interest" in learning course material; f) "effort" willing to expend during course. There were also two questions relating to experiences with prior, similar courses: g) "What is your average grade for previous course(s) in this particular subject only?"; and h) "overall, how effective was the instruction in your previous courses in this department?"

For the post-course questionnaire, questions a through f were rephrased (i through n) to find out about students' attitudes concerning the fulfillment of their expectations as a result of taking the course. Two additional questions asked about students' perceptions of the teacher's "enthusiasm" and ability to present "entertaining" lectures. Twenty-six questions were asked which pertain to students' assessment of the teacher's general ability or "skill", "rapport" with students, course "organization/structure, "workload" or difficutly, and "general course attitudes". That is, specific items (which are to be factor analyzed into five meaningful dimensions) were selected on the basis of their loading on one

of the five factors alone. The last five questions on the post-course questionnaire asked how important each of the following items were to students when rating the overall course: q) "exam fairness"; r) "rapport" with students' s) "entertaining" lectures, t) "amount learned"; u) the teacher's "ability to motivate" students' interest.

Analysis. Zero-order correlations examined the relationship between each background characteristic, pre-course motivation-related characteristic, and post-course motivation-related characteristic with both general teacher/course ratings and factor scores of specific ratings. A series of regression analyses was used to determine the successive predictive power of background, pre- and post-course motivation-related characteristics upon general items and specific factor ratings. In addition, two stepwise regression procedures attempted to determine which background characteristics best predicted pre-course motivation-related characteristics, and which post-course motivation-related characteristics along with the "enthusiasm" and "entertainment" variables would best predict general course ratings (after the effect of specific course ratings had been partialled out).

Results for Economics Sections

Zero-order correlations with general rating items. Table 1-1 (A, B) presents correlation coefficients representing the bivariate relationships between pre-course questionnaire items and general teacher/course ratings; background and pre-course motivation-related variables were correlated with students' rating of "overall teaching ability", teacher-student "rapport", course "content" coverage, and course "organization/ structure" for each of the four economics classes, E1, E2, E3 and E4. In addition, the ratings from pre-course questionnaire items were correlated with students' perceptions of the teacher's "enthusiasm" for the material and ability to present "entertaining" lectures, which were rated at semester's end. These correlations are based upon sample sizes of 102, 75, 85, and 90 for the four course sections, respectively. (The total enrollments are listed in Table 1-2.) The number of correlation coeffieients significant beyond the $p<.01$ level within each class section were 38, 31, 32 and 38, respectively.

Conspicuously absent were any significant zero-order correlations (in any class section) between background characteristics and either pre-course motivation-related

variables or end-of-semester general teacher/course ratings, with one exception. In section E3, student "major" correlated significantly ($r=.29$, $p<.01$) with rated teacher-student "rapport". However, the variance accounted for was less than nine percent as indicated by the r^2. All other individual bivariate relationships with background variables typically accounted for less than four percent of the variance between the questionnaire items and, more often than not, approached zero. Background variables infrequently produced large significant correlations when correlated with themselves. (For the sample sizes employed, the correlation was significant at $p \leq .01$ only if r was about .30 or greater.) The two exceptions consistent within each class section were obvious and uninteresting: "GPA" correlated with "GPA last semester only" ($.79 \leq r \leq .84$) and course "optionality" correlated with student "major" ($.44 \leq r \leq .68$). The most important result, then, pertaining to background variables was that they tended not to correlate significantly with either pre-course motivation-related variables or end-of-semester general teacher/course ratings.

Pre-course motivation-related variables correlated significantly with general teacher/course ratings on five different occasions. "Amount expected to learn"

correlated significantly ($p<.01$) with overall course "organization/structure" in section E2.
In addition, students' rated "interest" in the course correlated significantly with overall teacher "skill" ($r=.30$) and "rapport" ($r=.27$). The intercorrelations between pre-course motivation-related variables (items 9 thru 14, pre-questionnaire) were moderately strong ($.27\leq r\leq .51$) and, in general, highly significant ($p<.001$). The conclusion, with respect to students' pre-course motivation-related variables, is that they do not tend to account for much of the variance in end-of-semester general teacher/course ratings.

The general teacher/course (items 10 thru 13, post-questionnaire) ratings themselves were also highly intercorrelated ($.37\leq r\leq .75$). However, their relationship with students' perceptions of the teacher's "enthusiasm" and the "entertainment" value of lecture presentations (items 8 and 9, post-questionnaire) is more interesting, since both of these post-course variables correlated strongly ($.29\leq r\leq .69$) with each of the general teacher/course rating items. The largest correlations were obtained of overall teacher "skill" (GS) with both teacher "enthusiasm" (.61, .42, .58, .64) and with "entertainment" value of lectures (.51, .46, .69, .64). These results appear to be

similar to results reviewed earlier pertaining to the "Dr. Fox Effect" (Naftulin, et al., 1973; Ware, 1974; Ware and Williams, 1975; Granzin and Painter, 1973)

Other post-course ratings that produced large significant correlations ($.29 \leq r \leq .55$, $p < .01$) with the general teacher/course ratings were students' assessment of their "interest" in course material and whether or not the "amount learned" was as much as anticipated. To be more specific, "interest" correlated significantly ($p < .01$) with each of the general ratings, but more strongly with rated "content" coverage (GC) and course "organization/structure" (GX) ($.31 \leq r \leq .55$). Students' ratings of "amount learned" correlated significantly ($.29 \leq r \leq .54$, $p < .01$) with rated overall teacher "skill" (GS) and "content" coverage (GC) in each economics class. "Amount learned" also correlated significantly ($p < .01$) with course "organization/structure" (GX) in economicis section E2, E3, and E4: $\bar{r}_{GX} \cong .44$.

These results were not unexpected. One might expect students' perceptions of what they have learned from taking a course and how interesting the material was to effect their perceptions of the quality of the instruction/course. What seems more important is the fact that, in general, the identical pre-course ratings did

not relate to a significant or meaningful degree with teacher/course ratings: ratings are not biased much by these student characteristics.

Conspicuously absent in the results of the correlational analyses reported in Tables 1-1, C and D, were any effects upon the general teacher/course ratings associated with students' perceptions of the "relevance" of the course material, "effort" expended, appropriateness of the "work-load" required during the course or even "grade satis-faction". Later, these pre- and post-course motivation-related variables will be regressed upon each general rating item (criterion variables). This analysis should determine whether or not these variables will collectively account for large porportions of variance in the general ratings.

Just one issue remains to be reported concerning bi-variate relationships: the correlation between students' "reported grade in similar courses" within the same disci-pline (economics) and students' rating of the "overall effectiveness of instruction in similar courses". These two variables correlate very stongly in each of the economics sections (.79, .80, .83, .58). It appears that students' recollections of the quality of instruction in prior similar courses is somehow related to their

recollection of the grade they received. However, whether or not the grade a student received in such a course caused them to rate the course more or less favorably at that time is another issue; an issue unanswerable from the present data.

Zero-Order Correlations for Composite Items: Three types of composite items were considered in these data analyses: "complete" factor-scores (CFs); "incomplete" composite-scores (IFs); and equally weighted "incomplete" composite-scores (IC). The "complete" factor-score method (Guertin and Bailey, 1970) produces scores based upon all variables in the factor analysis. The "incomplete" methods yield factor-scores based upon only those variables which load highly upon a particular factor and are the methods most preferred by theoreticians (Nunnally, 1967; Guertin and Bailey, 1970). The complete factor-score method is viewed less favorably because it is more likely to produce scores which capitalize upon chance fluctations in the original variables. Even though such methods have been widely used by educational researchers (refer to studies listed in Appendix C), recent evidence suggests that ICs may produce better estimates of factor-scores than those with optimal weights (Wainer, 1976; Dawes and Corrigan, 1974; Trites and Sells, 1955). Consequently, each of the three types of composite variables were used in Pearson correlational analyses identical to those already

reported for general course ratings. These analyses will examine the strengths and weaknesses of each variable type. The variable type least contaminated by student background and pre-course characteristics will be used in later analyses. The variables selected for use in both the equally and optimally weighted "incomplete" cases were chosen as a result of a series of orthogonal factor analyses performed on eighteen of the specific rating items for each of the four economics sections. The resulting factor pattern matrices were examined by visual inspection to insure considerable similarity between the final solutions in each group. (A maximum likelihood test of the similarity/difference between these solutions was performed which suggested that the solutions were statistically dissimilar. However, in lieu of the observed similarity of the solutions and several problems which developed with Joreskog's maximum likelihood program, the decision was made to continue the analysis based upon the assumed similarity of the solutions. The reader is referred to the Addendum for a complete discussion of the maximum likelihood problem.) Since the solutions appeared to be similar, a combined group analysis (all four economics classes together) was performed. The results of this analysis was used to obtain the composite scores.

The varimax rotated orthogonal solution produced four factors ("skill", "rapport", "content", and "organization/ structure") which accounted for 39%, 9%, 7% and 7% of the variance, respectively. The items loading heavily (approximately .5, or larger) are listed under their respective factor headings in Table 1-3. Therefore, the "complete" factor scores (standardized) were obtained using all the variables listed in the tables and the procedures specified in SPSS (Nie, et al., 1975). These procedures for calculating factor-scores are, incidentally, identical to those used in BMD (Dixon, 1974), according to Guertin and Bailey's (1970) documentation of the BioMed Program. The "incomplete" scores were linear combinations of disjoint subsets of variables (standardized) as specified in Table 1-3. The difference between "incomplete" factor-scores and "incomplete" composite scores is that the linear combination for the latter multiplies the observed variables by one rather than their respective factor-score coefficients. The results of the zero-order correlations between each set of these derived scores and background and motivation-related variables were quite interesting. (The reader should refer to Table 1-4 in order to familiarize him/herself with the notation and item numbers corresponding to the general and composite items and their meanings.)

Table 1-3: Items Loading on Factors

I. Skill

1. Class presentations were well-organized. (.76)*
2. The instructor was well-prepared for lectures. (.78)
3. The course material appeared to be presented in logical content units. (.57)
4. The instructor clearly specified course objectives. (.51)
5. The instructor expressed ideas clearly. (.70)
6. The instructor's voice was clear and understandable. (.48)

II. Rapport

1. The instructor used student questions as a source for discovering points of confusion. (.63)
2. The student was never hesitant to ask questions in this course. (.59)
3. The instructor encouraged students to ask questions in this course. (.68)
4. The instructor showed a genuine interest in teaching this course. (.35)

III. Content

1. The student feels he/she profited from out-of-class assignments. (.45)

Table 1-3: Items Loading on Factors (Continued)

2. This course was worthwhile because the content is directly applicable to the student's planned occupation. (.58)
3. The student would take another course with similar content even if he/she did not have to. (.64)
4. Homework assignments were interesting and stimulating. (.52)
5. Because of this couse the student developed an increased appreciation for the subject area. (.63)

IV. Organization/Structure (Exams)

1. Exams stressed conceptual understanding. (.57)
2. The grading procedures fairly indicated each student's performance. (.59)
3. Exams adequately covered the text material. (.71)

*Loading of each variable upon the specified factor for the orthogonal, varimax-rotated solution.

Table 1-4

		Skill GS	Rapport GR	Content GC	Organization/ Structure GX
(GI)	General Item				
(IC)	Incomplete Composite	S1	R1	C1	X1
(IFS)	Incomplete Factor-Score	S2	R2	C2	X2
(CFS)	Complete Factor-Score	S3	R3	C3	X3

In spite of the differences in methodology used to calculate the composite and factor score variables, the intercorrelations between the derived scores with identical meaning were all greater than .90 (e.g., the average correlates between "skill" items, S1, S2, S3, were $r_{S1,S2} \cong .98$, $r_{S2,S3} \cong .90$, $r_{S2,\ S3} \cong .93$, respectively). Therefore, one might expect the results observed for zero-order correlations between any "skill", "rapport", "content", and "organization/ structure" item and a background or motivation-related variable would be consistent over each of the composite/factor score items with similar meaning. (Therefore, to simplify the reporting of results, only data for composite scores (ICs) will be discussed. The relationships for the other derived ratings, incomplete and complete factor scores, are virtually identical to those reported for composite scores. However, the data for all three derived scores is represented in the data tables.)

To begin with, background variables (refer to Table 1-5 A through D) did not correlate significantly ($p<.01$) with any of the composite (factor-score) items. There was only one exception to this rule. The correlateion between "GPA" (grade point average) and organization/ structure (X3) was just significant ($r_{E1}=-.25$, $p<.01$). This single, atypical result, is not sufficient to conclude that specific (composite/factor-score) ratings are not generalization over student background variables. These results are consistent with the data reported for general teacher/course ratings.

Pre-course motivation-related variables, on the other hand, correlated significantly and substantially with the composite (factor-score) items. This result is at variance with that reported earlier for general rating items, both in number of significant correlations and in magnitude of the correlations. In general, pre-course motivation-related variables correlated a bit more strongly with the composite (factor-score) items ($.31 \leq r_S \leq .50$). The reader is again asked to refer to Tables 1-5 A through D.

Student - rated pre-course "interest" correlated consistently and significantly ($.30 \leq r \leq .46$, $p<.01$) with rated course content (C1) in each section of economics. In addition, rated pre-course "interest" was significantly correlated ($.29 \leq r \leq .37$, $p<.01$) with rated course "organization/

structure" (X1) in section E2 and E4. "Interest" also correlated significantly ($r = .31$, $p < .01$) with "skill" (S1) in section E4. All of these results are inconsistent with those reported for general rating items. That is, pre-course "interest" did not correlate significantly with any general rating items.

Rated pre-course "relevance" correlated significantly ($.28 \leq r \leq .35$, $p < .01$) with course "content" (C1) in sections E3 and E4. Rated pre-course "amount expected to learn" was also significantly related ($.27 \leq r \leq .52$) to students' ratings of the grade they would be satisfied with at semester's end ("grade satisfaction") and also correlated significantly ($.37 \leq r \leq .39$) with rated teacher "skill" (S1) in section E4.

These results suggest that pre-course motivation-related variables, particularly students' "interest" in the course and students' perceptions of the "relevance" of the course material, can in certain instances influence end-of-semester ratings. Those composite (factor-score) items most affected by student attitudes seemed to be rated course "content" and course "organization/structure". The variance accounted for between variables in these relationships was observed to be as much as twenty-five percent. These results tend to refute one hypothesis proposed in this research: that specific rating items

or variables composed of some combination of specific rating items would be less likely to be contaminated by students' pre-course attitudes. Students' ratings of both "skill" and "rapport" tended to be free of bias that might result from such pre-course attitudes, regardless of whether the rating items were general or specific in composition.

Post-course motivation-related variables did correlate significantly with composite (factor-score) ratings, but not as consistently or frequently as reported for general items. The post-course items possessing both consistency and frequency of occurance are rated teacher "enthusiasm", "entertainment" value of lectures, "interest" in the course material, "relevance" of the course material, and the "amount learned" by the students (refer to Table 1-5 A through D).

Typically, teacher "enthusiasm and "entertainment" correlated significantly with "skill" (S1) and "rapport" (R1) for each of the four economics courses ($.22 \leq r \leq .54$, $p < .01$), although the magnitude of the correlations seemed slightly smaller than those reported earlier for general items (GS, "skill" and GR, "rapport"). While it was true that the two remaining general items (GC and GX) also correlated significantly with rated "enthusiasm" and "entertainment", this was not found to be true for composite (factor-score) items. Students' "interest" in

the course material tended to correlate significantly with each of the composite (factor-score) items, however, the proportion of variance accounted for was most often less than ten percent. One exception was "content" coverage (C1) in each course section which correlated at least .55 and typically more with "interest" as indicated by the average correlation, $\bar{r}_{C1} \cong .63$. Note that the average correlation representing the same relationships between post-course items and the same general rating item (GC, "content") was much smaller, $\bar{r}_{GC} \cong .40$.

Rated "relevance" of the course material also correlated significantly ($.37 \leq r \leq .60$, $p < .01$) with course "content" (C1) in each economics class. The average correlation was $\bar{r}_{C1} \cong .49$. "Relevance" did not correlate significantly with any of the other general teacher course rating items, a result contrary to that just reported for the composite items.

The only other relationship observed which was consistent over the four economics groups was students' ratings of "amount learned" and "skill" (S1). Even though the correlations weresignificant ($.22 \leq r \leq .46$, $p < .01$), they were usually small as indicated by the average correlation, $\bar{r}_{S1} \cong .40$. The results obtained for S1 were consistent with those already reported for the general rating item ($\bar{r}_{GS} = .43$).

In sum then, these results suggest that general rating items (GS, GR, GC and GX) tend to be more strongly related to students' post-course assessment of teacher "enthusiasm" and "entertainment" value of the lectures than to their composite (factor-score) counterparts (e.g., S1, R1, C1 and X1). However, this trend seemed to be reversed for the relationship of "interest" with "content" (GC). The general item relationships were small and non-significant, while large and highly significant for the composite (factor-score) items with similar meaning.

Before proceeding to a description of the results of the several regression analyses, a few more bivariate relationships should be mentioned briefly. In particular, the intercorrelations between the composite, factor-score, and general ratings themselves.

The extent of the intercorrelations between each of the items S1 through X1 has already been described. To recapitulate, the correlations between items of similar meaning approached unity (refer to Table 1-6 A and B). However, the correlations between general ratings (GS, GR, GC, and GX) and composite (factor-score) items (S1 to X1) were less substantial than anticipated ($.18 \leq r \leq .76$). However, the correlations between the variables of interest (e.g., "skill", GS with S1,

"rapport", GR with R1, "content", GC with CI and "organization/structure", CX with X1) were all moderately large ($.33 \leq r \leq .76$). In effect, general items with meanings similar to composite (factor-score) rating items did not tend to measure exactly the same characteristics of teachers/courses. Evidence to support this statement is obviated by the observed range in proportion of variance accounted for by these relationships ($.11 \leq r^2 \leq .58$). There are undoubtedly a variety of reasons why this might be true. One likely explanation might be that the composite items do not represent all of the specific criteria strudents use when making their general assessments of teachers/courses. Some of these criteria such as teacher "enthusiasm" or "entertainment" value of lectures might be unrelated to traditional measures of academic performance.

In the paragraphs which follow, the results of several multiple regression analyses will be summarized, adding some credibility to the last few statements. Broadly speaking, the regression analyses will help determine the extent to which general teacher/course rating items can be predicted by knowing students' background and course motivation-related variables. These empirical results will then be compared to results obtained from the analogous analyses, where the general items have been

replaced by an appropriate composite (factor-score) item.

Regression analyses and student background, pre-course and post-course motivation-related characteristics. The analyses were performed in a hierarchical manner, with groups of variables being entered into the regression equation based upon their natural temporal order of occurance. That is, background variables were hypothesized to be casually linked to pre-course motivation-related variables. Post-course motivation-related variables, on the other hand, were considered to be a simple reassessment by students of the pre-course variables at semester's end. Therefore, the order of entry of these sets of variables into the regression equations was as follows: background variables (B), pre-course motivation-related variables (P), and post-course motivation-related variables (P).

The results of the regression analyses predicting each of the general rating items for each of the four economics courses produced several interesting results. In no instance did a linear combination of background variables ("GPA", "college year", "major", "optionality", and preferred "class size") account for a significant proportion of variance for any of the general items (refer to Table 1-7A). In most instances, the variance accounted for was approximately five percent. Similarly, the proportion of variance accounted for by the combination of both students' background and students' pre-course motivation-

related variables ("amount expected to learn", "relevance", of course material, the desirability of the course "workload", "interest" in taking the course, and "effort" willing to expend for the course) were typically non-significant with the exception of two equations. The proportion of variance accounted for , $R^2_{y \cdot PB} = .30$, ($F(10,57) = 2.48$, $p<.05$) by the linear combination of the two sets of variables in section E2 was significant in predicting course "organization/structure". In addition, in Section E3 the same two sets of variables accounted for twenty-five percent ($R^2_{y \cdot PB} = .25$, $F(10, 68) = 2.16$, $p<.05$). However, it is a well known fact that multiple regression analysis over-estimates the proportion of variances accounted for because of chance variation in the zero-order correlations between predictors and criterion variables (Cohen and Cohen, 1975, p. 106). Therefore, it is recommended that one consider the adjusted $R^2 = \tilde{R}^2$ which compensates for such over estimations. The correction formula recommonded for estimating R^2 is provided by Cohen and Cohen (1975, p. 106):

$$\tilde{R}^2 = 1-(1-R^2)\cdot(n-1/n-k-1),$$

where n = total number of subjects; k = number of predictor variables; and R^2 = observed squared multiple correlations.

The adjusted R^2(s)($\tilde{R}^2_{E2} = .20$, $\tilde{R}^2_{E3} = .15$) for the relationships of rated course "organization/structure" and

rated "rapport" with pre-course student variables (B and P together) were not significant. However, the adjusted proportion of variance accounted for by the pre-course motivation-related variables (P) alone when predicting rated course "organization/structure" in section E2 was significant ($\tilde{R}^2_{y \cdot PB} - \tilde{R}^2_{y \cdot B} = .20$, $F(5,57) = 3.25$, $p<.05$). This result should be viewed cautiously, though, since the experiment-wise error would be approximately $1-(1-.05)^{16} = .44$. Therefore, one might expect several spurious significant results due to chance alone. To compensate for such experiment-wise error rates, one could divide the alpha-level by the total number of parameters being tested: .05/16 = .003 (Cohen and Cohen, 1975, p. 156). The error level then becomes a great deal more conservative for individual F-ratios. If the one significant result reported previously is viewed in this light, it does not seem meaningful. Consequently, one would like to argue that neither student background variables nor student pre-course motivation-related variables are significantly related to general teacher/course ratings. These findings are the opposite of what was expected.

The influence of post-course motivation-related variables upon general teacher/course ratings was substantial. Eight out of sixteen of the adjusted ($\tilde{R}^2_{y \cdot \underline{P}PB} - \tilde{R}^2_{y \cdot PB}$)

linear combinations of these post-course variables accounted for significant ($p<.05$) proportions of variance in the general ratings. The larger and more consistent effects were in predicting ratings of course "content" (GC) and course "organization/structure" (GX), the largest of which accounted for thirty-four percent of the variance in the criterion variable. Furthermore, the combined effects of the three sets of variables ($\underline{P}$, P, and B) were significant in 10/16 instances accounting for as much as forty -five percent of the variance in the criterion variable. These two results taken together suggest that students' post-course motivation-related attitudes are related to teacher/course ratings. However, the magnitude of the effects of these variables is not dramatic, as indicated by the adjusted squared, semi-partial correlation coefficients ($.16 \leq \tilde{R}^2_{y.\underline{P}PB} - \tilde{R}^2_{y.PB} \leq .36$). About half of the squared semi-partial correlations were less than .16. These latter results were anticipated, since one would expect that students' perceptions of post-course items, like course relevance", "interest" in the course, and "amount learned" by the students, would be affected by their participation in and undergoing of a semester of course activity.

Another equally interesting question is whether the results reported for general rating items were also obtained for either composite or factor-score variables. The over-

all pattern of results observed from predicting these derived criterion variables (IC(s), IFs(s), and CFs(s), as stated in Table 1-4) from students' pre-course characteristics was very similar to that already reported for the general items, although the amount of variance accounted for in the course "content" (C1) relationships was substantially larger (refer to Tables 1-7B, C, and D). The major results were that neither background nor pre-course motivation-related variables accounted for a significant portion of the variance in the criterion variables with the exception of students' ratings of course "content". In three instances (sections E1, E3, and E4) the pre-course motivation-related variables accounted for significant ($p<.05$) proportions of variance in the criterion variable of course "content" ($.16 \leq \tilde{R}^2 \leq .22$). In addition, the effects of post-course motivation-related variables upon the composite (factor-score) criterion variables were quite similar to those already reported for general ratings with the exception again being ratings of course "content". In each of the four course section, the variance accounted for was significant beyond the .01 level ($.25 \leq \tilde{R}^2 \leq .45$). The variance accounted for in these relationships is substantially greater than that accounted for by the general rating items: .25, .02, .17, and .34, respectively. In

addition, it should be noted from looking at Table 1-8 that both rated instructor "skill" and rated course "organization/structure" are somewhat predictable by students' post-course attitudes.

To summarize, students' background variables do not seem to influence ratings of teacher "skill", "rapport", course "content", or course "organization/structure", whether the ratings are general or specific in nature. Pre- and post-course ratings do predict to a considerable extent students' ratings of course "content" for composite/factor score rating items. This result is not in discord with the hypothesis stated earlier, since post-course attitudes were expected to be influenced by the semesters' course activity and, therefore, correlations with students' ratings of the teacher/course.

It should be added in passing that pre-course motivation-related variables (e.g., "interest", "relevance", "amount expected to learn", etc.) are not predictable to any extent by student background variables. However, the results reported above do not suggest that the pre-course items are of any more utility in predicting end-of- semester teacher/course ratings than background variables. However, post-course motivation-related variables are very useful as predictors of teacher/course ratings at semester's end as the data suggest. Students who learned more, who

perceived the material as "relevant" to their needs, and who found the material of "interest" tend to rate the teacher/course more highly, although these items tend to be more useful in predicting students' ratings of course "content". These results are intuitively obvious.

Regression analyses and post-course motivation-related variables. The purpose of this analysis was to determine how much variance would be accounted for by post-course motivation-related variables when predicting the general rating items after the variance accounted for due to the composite (factor-score) items had been partialled out. Stated differently, after removing the variance due to performance criteria, will post-course motivation-related characteristics further predict general ratings? As noted earlier, the intercorrelations between the derived variables with similar meaning approached unity (e.g., $r_{S1, S2}$=.97, $r_{R1,R2}$= .98 etc.). Therefore, the results discussed in this section will be reported only for the equally weighted "incomplete" composite scores. That is, the results obtained using any one of the three sets of derived variables produce identical results when predicting the general ratings. The empirical findings here are completely generalizable to each of the three sets of ratings: ICs, IFs, and CFs.

The composite rating items (S1 = "skill", R1 ="rapport", C1 = "content", and X1 = "organization/structure") consistently accounted for large significant ($p \leq .01$) proportions of variance in each of the genral ratings for each economics section ($.30 \leq \tilde{R}^2 \leq .61$). The second set of variables (post-course motivation-related) entered into the equations did not predict a significant portion of variance for any of the general ratings. This result is somewhat surprising since this set of variables consistently predicted large portions of the variance in the general items in the first set of regression equations. Apparently, the attitudes being measured in the post-course motivation-related variables is redundant with the information already contained in the composite variables. The remaining two predictor variables (teacher "enthusiasm" and "entertainment" value of lectures) proved to be strongly related to the criterion variables of "skill" ($.10 \leq \tilde{R}^2 \leq .15$) in sections E1, E3, and E4, and "rapport" ($.04 \leq \tilde{R}^2 \leq .09$) in each of the four sections. Even though the results are quite consistent, the magnitude of the relations were not overwhelming. These results were anticipated, although, expected to be larger in magnitude. These data suggest two things: fist, that the post-course motivation-related variables used in the research must be

related to the composite rating items of "skill", "rapport", "content", and "organization/structure", and, second, that the general ratings of "skill" and "rapport" are dependent upon students' assessment of the teacher's "enthusiasm" and/or the "entertainment" value of his/her lectures.

These findings seem to be somewhat consistent with those reported by Naftulin, Ware, and Donnelly (1973) and Ware and Williams (1976). However, the magnitude of the relationships in these data are weak enough to preclude making any dramatic statement about the effects of either "enthusiasm" or "entertainment" upon teacher/course ratings. The data do suggest that both teacher "entertainment" and teacher "enthusiasm" are important to students when making their teacher/course ratings but not as important as Naftulin, et al., suggested in this statement: "Faculty who master the Dr. Fox effect may receive favorable student ratings regardless of how well they know the subjects and regardless of how much students learn" (1973, p. 12). The problem here, as in their study, is that circumstance beyond the control of the experimenter prevented the recording of the appropriate achievement data which might have shed some more light upon the relationship of "enthusiasm",

"entertainment", and the actual "amount learned" by the students with their ratings of college teachers/courses.

It should be added that when the general ratings are used in regression equations to predict composite variables (refer to Table 1-11A, B, and C), the effects of teacher "enthusiasm" and "entertainment" disappear: they have zero predictive power of the composite items. However, rather interesting strong results were obtained from post-course motivation-related variables. In particular, these post-course items accounted for large portions of variance in students' ratings of course "content" ($.21 \leq (\tilde{R}^2_{y \cdot \underline{P}G} - \tilde{R}^2_{y \cdot G}) \leq .48$) and to a lesser extent, ratings of course "organization/structure" ($.07 \leq (\tilde{R}^2_{y \cdot \underline{P}G} - \tilde{R}^2_{y \cdot G}) \leq .32$) in each course section. Each of the "delta" $\tilde{R}^2$(s) was significant beyond the $p<.01$ level. It appears that post-course motivation-related variables are strongly related to students' ratings of course "content", accounting for variance over and above the variance accounted for by the general ratings themselves ($.08 \leq \tilde{R}^2_{y \cdot G} \leq .40$). Also noteworthy was the tendency for general rating to consistently predict less variance in the "complete" factor-score items as compared to the other two composite measures. This can probably be attributed to the increased variance associated with these factor-score ratings as a result of the use of

all of the specific items to represent each factor. In such a case, one would suspect that the scores would contain considerable variance unrelated to the dimension in question. In effect, using all the items to derive scores for each factor would probably result in scores with more total and, therefore, more error variance than that attributed to "incomplete" scores.

A brief outline and summary of the major results of this section is as follows:

1. Background variables ("GPA", "college year", major", "optionality", "class size") did not correlate significantly with either the general or specific items ("skill", "rapport", "content", "organization/structure") in any of the economics sections.

2. While pre-course motivation-related variables ("amount expected to learn", "relevance", "workload", "interest", "effort") did not correlate significantly with general ratings of the teacher/course, several of them did correlate significantly ($.27 \leq r \leq .46$, $p < .01$) with composite (factor-score) rating items. In particular, "amount expected to learn", "relevance" of course material, and "interest" in the course were highly correlated with rated course "content" coverage. However, "interest" was the only pre-course motivation-related variable correlated with any rating item consistently over each economics course section (E1 through E4). This last result was contrary

to those predicted to occur. Composite (factor-score) items were expected to be free of bias from pre-course items.

3. Post-course motivation-related characteristics (identical in meaning to those variables listed in #2 above) correlated significantly with both general and composite (factor-score) items in each economics section. Of particular interest was the fact that those correlations seemed considerably larger for the composite factor-score) items. Again, "interest" correlated with "content" and produced the larger correlation coefficients. The observed relationships for composite (factor-score) items were all greater than or equal to .55. Course "relevance" produce the second largest group of correlations with the "content" rating items in each of the sections E1 through E4. Even though "relevance" was not related to the general rating items, it was highly correlated with course "content" for composite (factor-score) items ($.37 \leq r \leq .60$).

4. Rated teacher "enthusiasm" and "entertainment" value of lectures were highly correlated with each of the general and composite (factor-score)rating items. However, the effects of these two variables, "entertainment" and "enthusiasm", seemed to be more strongly related to general ($.42 \leq r \leq .69$) than composite (factor-

score) items ($.22 \leq r \leq .54$). These observed results were hypothesizedearlier, although the magnitude of the relationships is somewhat larger than anticipated for the composite (factor-score) items.

5. Post-course assessments of "workload" desirability and "effort" expended were unrelated to any of the teacher/course rating items.

6. As one might have suspected based upon the zero-order correlational relationships reported above, background variables did not predict (multiple regression analysis) teacher/course ratings to any meaningful or significant extent. The pre-course ratings did not predict general ratings, but did predict ($.16 \leq \tilde{R}^2_{y.PB} - \tilde{R}^2_{y.B} \leq .22$) "content" composite (factor-score) items. This result was not expected. Composite scores of specific rating items were expected to be free from students' pre-course attitudes (McKeachie, Lin, and Manning, 1971).

7. Students' attitudes were observed to have changed as indicated by the high correlations of the post-course variables with teacher/course ratings. Linear combinations of these post-course variables accounted for large portions of the variance in both general items ($.16 \leq \tilde{R}^2_{y.\underline{P}PB} - \tilde{R}^2_{y.PB} \leq .45$) over and above that accounted for by the corresponding pre-course variables. In addition, it appears that composite (factor-score) items are more affected by post-course

variables than their general rating item counterparts. These results were also expected, as stated in the hypotheses at the beginning of this section.

8. The results of the last set of regression analyses demonstrated that each type of rating item (general or composite factor-score) was highly predictable from linear combinations of the other type. For example, a large portion of the variance in general teacher "skill" ($.30 \leq \tilde{R}^2_{y \cdot C} \leq .59$) was accounted for by the composite items (S1, R1, C1, and X1). Similarly, the variance in S1 was largely accounted for by GS, GR, GC, and GX ($.50 \leq \tilde{R}^2_{y \cdot G} \leq .69$), but to a greater extent than that observed for the general items. Perhaps more interesting was the observed relationship for sets of post-course variables with the general and composite (factor-score) criterion variables. Post-course motivation-related variables did not account for any variance in general items over and above that already accounted for by the set of composite items. However, in comparable analyses predicting the composite scores, the post-course variables accounted for a considerable portion of variance ($.21 \leq \tilde{R}^2_{y \cdot PG} - \tilde{R}^2_{y \cdot G} \leq .48$) in course "content" and "organization/structure" items, in addition to that already reported for the set of general items. When "enthusiasm" and "entertainment" are added to these

equations, the two variables predict general items ("skill" and "rapport"), but do not predict any of the composite (factor-score) items. The variances accounted for in the two significant relationships with general items were $.10 \leq \tilde{R}^2_{y \cdot I\underline{P}C} - \tilde{R}^2_{y \cdot \underline{P}C} \leq .15$ and $.04 \leq \tilde{R}^2_{y \cdot I\underline{P}C} - \tilde{R}^2_{y \cdot \underline{P}C} \leq .09$, respectively. Clearly, the "skill" variable is more strongly related to a teacher's "enthusiasm" and the "entertainment" value of his/her lecture presentations.

Discussion

What conclusions concerning the generalizability and, therefore, the utility of students' ratings of college teachers and courses, can be derived from the results of the present research? Which of the two rating schemes (general or specific items) appears to be more desirable from the generalizability perspective? That is, if the data do in fact suggest that a clear choice can be made between either type of rating item based upon the empirical findings of this research, then which should be chosen and why. Unfortunately, a clear choice between the rating schemes probably cannot be made. Each type of item has measurement problems associated with it: in effect, the problems concern the generalizability of the ratings. The final choice

between rating schemes will have to depend upon the purpose of the evaluation.

The generalizability of general and composite rating items: Zero-order correlations. "The important feature of generalizability is that it directs attention to any differences due to the people whose opinions and observations are requested, differences due to the occasions of teaching that are evaluated, differences due to the circumstances surrounding the evaluation and so forth" (Doyle, 1975, p. 71). It was to this purpose that the current research was directed and upon which this discussion will focus.

Student background variables ("GPA", "college year", "major", "class size", and "optionality") were not related to either general or composite ratings of teachers/courses. General and composite ratings of teacher "skill", teacher-student "rapport", course "content", and course "organization/structure" can be sampled without regard for these student background variables. The only surprising result was the lack of relationship between course optionality and any of the teacher/course ratings. Evidence reviewed earlier indicated that "GPA", "college year", "major", and "class size" typically have little or no relationship to teacher/course ratings (Costin, et al., 1971; Kulik and

Kulik, 1974; Aleamoni, 1974, to mention a few). However "optionality" was contently reported to be negatively related to teacher/course ratings (Costin, et al., 1971; Doyle and Whitely, 1974; Granzin and Painter, 1973; Kulik and Kulik, 1974; Kulik and McKeachie, 1973). The inconsistency between the reviewed literature and the empirical findings of this research could be a result of dichotomizing the issue of "optionality": elected or required course. The dichotomy does not clearly separate the reasons why many students may be enrolled in a particular course. For example, some students may have wanted to take the course even though it was required. Alternatively, other students may have elected the course for reasons of expediency, such as it fitting into their schedules. Therefore, the relationship of course "optionality" to teacher/course ratings might be better assessed if the "optionality" question were presented to students with many different response alternatives. However, even this approach has problems since one is unlikely to tap all the alternatives. Perhaps a better solution to the probem would be to simply ask students, "How much did you want to take this course?" There is currently no empirical literature available using this type of question to measure the relationship of course "optionality" and teacher/course ratings.

Pre-coure motivation-related student variables did relate substantially with composite ratings of course "content" and "organization/structure". General ratings of "skill", "rapport", "content", and "organization/structure" as well as composite ratings of "skill" and "rapport" were generalizable over pre-course motivation-related variables, in particular, "amount expect to learn" by the students, perceived "relevance" of the course material to planned career, desired difficulty of the course "workload", expressed "interest" in the course, and "effort" willing to invest in the course. This result is interesting since it suggests that students come into a new course without allowing past experience to influence their opinions of the teacher's classroom performance. However, students' opinions about specific aspects of the course "content" and "organization/structure" seem to be systematically affected by other experience. Presumably, these prior experiences related to academic course work, although the current data do not make provisions which would allow a conclusive answer concerning this issue. The most consistent influence upon the rated course "content" and "orgranization/structure" was by far students' rated "interest" in the course. It seems somewhat unusual that composite items would reflect

student "interest" and not the general ratings of similar meaning.

A possible reason why the general and composite ratings were differentially affected could be that the rating items are measuring different aspects of the "content" and "organization/structure" of the courses. These differences in measurement are reflected in the intercorrelations between the general and composite "content" rating items (.31, .44, .50, .156,respectively). However, the exact differences between the two variable types cannot be determined from the present data. An approximation can be obtained from examining the zero-order correlations between the specific items and the appropriate general items.

The relationship of students' pre-course "interest" in the course material may be a result of their stated liking for the discipline (i.e., the social sciences, item #18, pre-course questionnaire). It is not a result of the grade received or the quality of instruction in a prior similar course. Furthermore, the students' "interest" was not related to "GPA", "college year", "major", "optionality", or "class size". These facts were clearly demonstrated by a series of regression analyses (Table 1-9).

In one instance, rated teacher "skill" (composite score) was associated with students' ratings of "grade satisfaction", the grade that they would be satisfied with

receiving (r = .37). Because this occurred in only one course section it might be attributed to chance variation. One could argue that such a result should be expected, since the literature reviewed (e.g., Costin, et al., 1971; Kulik and Kulik, 1974; etc.) suggests that students' grades typically correlate (r = .30) with rated teacher performance. In this particular case, students were merely expressing a "wish" (most students would not be satisfied with less than an A or a B; see Appendix A). However, if one looks at the correlations between "expected grade" and that "wished" for, it appears that most students' dreams came true (r = .60).

Unfortunately, in this research, no absolute conclusion can be reached concerning student grades and teacher ratings, since one faculty member objected to providing this information. As a result, each of his three colleagues agreed to follow his example and decided not to provide the data either.

In sum, certain pre-course motivation-related variables are strongly related to composite teacher/course ratings. And in the case of course "content", the more influential pre-course ratings were "amount expected to learn", perceived course "relevance", and most important, students' "interest" in the course. The importance of these items was determined by considering their standardized beta weights. Even though the beta weights were not entered in

the optimal ordering, the magnitude of these three items ("amount expect to learn", "relevance", and "interest") was consistently much larger than those for "workload" and "effort". In each case, the more students expected to learn, the more relevant the course material; and the more interesting the course material, the more positive the rated "content" and "organization/structure". It can be aruged then, that certain pre-course attitudes do tend to contaminate composite "content" and "organization/structure" ratings. Alternatively, it might be that these pre-course ratings are based upon course descriptions of the intended content coverage which are available to students. In this case, the composite ratings may be reflecting the consistency of the actual course content with the pre-course description. In any case, students may either like or dislike the material covered.

In several instances, both general and composite teacher/course ratings were affected by post-course student variables. The students' ratings of "amount learned", "interest" in the course, teacher "enthusiasm", and level of "entertainment" in lecture presentations were strongly related to rated teacher "skill", "rapport", "content", and "organization/structure". In addition, rated "relevance" of the course material and "effort"

expended by students correlated sporadically with rated composite "content" and other teacher/course rating items. However, the inconsistency of these results render them somewhat less interesting and unimportant. The influence of post-course motivation-related variables are not viewed as contaminants of teacher/course ratings. On the contrary, they reflect student attitudes which have changed or were affected as aresult of having participated in the course. However, the obvious effects of teacher "enthusiasm" and "entertainment" level are more often than not perceived by factuly members and researchers (Costin, et al., 1971; Kulik and McKeachie, 1973; Naftulin, et al., 1973; Ware and Williams, 1975) as undesirable. The fact that teacher/course ratings seem to depend upon both the "amount learned" by students, the teacher's "enthusiasm", and the teacher's ability to "entertain" seems to support the notion. It would be interesting to determine which of these three items are most important to students when making their ratings of a teacher's general "skill". This could perhaps be achieved by doing some type of path analysis: but this is somewhat beyond the scope and purpose of the present research.

In pointing out that "amount learned", "interest", and"enthusiasm" were strongly related to teacher/course ratings, it should also be noted that these three charac-

teristics have been investigated frequently in the past. The teacher's ability to get students interested in the course material, the teacher's enthusiasm for the subject matter, as well as his/her ability to motivate students to do their best work have been described by both Costin, et al. (1971) and French-Lazovik (1974) as the characteristics most often cited in effective teachers. Effective teachers were described as: 1) possessing a thorough knowledge of the subject matter; 2) having the ability to explain clearly and use examples well; 3) being enthusiastic, energetic and having a lively interest in teaching; 4) being student-oriented and friendly; 5) being well prepared and organized for lectures; 6) able to get students interested in the subject matter; and 7) motivating students to do their best work.

Should the "amount learned" by students and students' "interest" in the course material be considered contaminants of teacher/course ratings? Probably not. It would seem desirable to have teacher/course ratings which reflect students' opinions of these course-related attitudes since the individual faculty member has a good deal of control over students' opinion of these two variables. In every case, the more students determined that they had learned and the more interest they had in course material, the more positive were the teacher/course ratings.

Similar arguments might be offered in support of the effect of teacher "enthusiasm" and the "entertainment" level of lecture presentations, especially if these items could be shown to be casually linked to student learning. Each of these items reflect the teacher's individual character and personality and should effect, to some extent, students' opinions of the teacher and course. Considerable evidence has been offered to support this argument on behalf of teacher "enthusiasm" (Costin, et al., 1971; French-Lazovik, 1974; Costin and Grush, 1974; Grush and Costin, 1975; and Sherman and Blackburn, 1975). Support for the influence of teacher "entertainment" is a bit more difficult to marshal.

The effects of teacher "entertainment" have been portrayed, for the most part, in an unfavorable light. In particular, Naftulin, et al. (1973) and Ware and Williams (1975) have suggested that faculty who master the "Dr. Fox Effect" (teacher reliance on charisma and entertainment rather than provision of substantive information) would receive favorable ratings even though students had learned very little. Unfortunately, these researchers either had no learning measure at all or failed to achieve an efficient coherent measure methodologically. To a limited degree, ratings do become more favorable as the "entertainment" value of the lectures increases, as demonstrated

in the present research and in some earlier research by Granzin and Painter (1973). However, one should not go so far as to suggest that in the absence of learned material students would still rate a course favorably on the basis of "entertainment" alone. Clearly, the "amount learned" does play a part in students' ratings of teachers/courses.

The evidence presented here does not resolve the controversy over the effects of teacher "entertainment" upon teacher/course ratings. However, if an achievement measurement had been obtained in the present research, then the casual effects of the actual "amount learned" by students, student "interest", teacher "enthusiasm", and the "entertainment" value of lectures could have been investigated. The utility of path analysis in resolving these types of issues is explained in Nie, et al. (1975), Cohen and Cohen 1975), and Kerlinger and Pedhazur (1973).

Interestingly, the relationships of teacher "enthusiasm" and teacher "entertainment" were strongest for general teacher items, particularly the "skill" variable. Therefore, until it is conclusively shown that teacher "enthusiasm" and teacher "entertainment" induce higher levels of student learning, one can only conclude that general ratings of "skill" are sufficiently biased by these variables to sug-

gest that a composite variable to be used to assess teacher competence. This position would be of particular importance if the data were being used in personnel decisions.

The "effort" students expended during the course was also related to composite ratings of course "content" but not to general ratings, which seems reasonable if one considers the items composing the course "content" ratings. Several of the items pertain to "content" coverage of the out-of-class assignments (e.g., "Homework assignments were interesting and stimulating", and "I feel I profitted from out-of-class assignments"). In effect, the more effort students expended on the course, the more positive were their ratings of course "content". This is understandable, since the more effort students invest working on course material, the more worth they would find in it. And because of the amount of effort expended on the course, they would, presumably, be better able to make a critical assessment of content coverage. The differential effect the "effort" variable had upon the general and composite items probably reflects, once again, the fact that these rating items are not measuring identical facets of the course.

Regression analyses. The results of the first set of hierarchical regression analyses would have been more interesting had the sample sizes been larger. As it turned out, the natural selection or drop-out rate of subjects due to students not completing one of the question-

naires was relatively high (as much as sixty-seven percent). However, one can assume that the drop-out rate was the result of a random process, since the students had no way of knowing when the evaluation questionnaires would be administered. In addition, if the proportion of missing data is very large, as with these data, Cohen and Cohen (1975) have argued convincingly that it makes little sense to try to include the missing data in the analyses, that is, by recoding variables; for example, plugging with variable means (refer to p. 286). The more serious problem which arises because of missing data is the associated loss of statistical power necessary to test analytic results. The reader should keep these comments in mind when reading the paragraphs which follow pertaining to the results of the regression analyses. In many instances, large portions of variance were accounted for by the analyses, but failed to reach significance because of the reduction in power.

Both general and composite teacher/course ratings were predicted from background, pre-course and post-course student variables. The adjusted (shrunken) semi-partial correlation coefficients for the three sets of predictor variables accounted for a substantial proportion of variance in each of the general and composite teacher/course rating items. The effects were most pronounced

when predicting the "content" ratings and particularly for composite variables. The major portion of the variance accounted for (also most consistent) in these relationships was due to the predictive power of the post-course motivation-related variables ("amount learned", "relevance", "workload", "interest", and "effort"). These findings are not surprising when the magnitude of the relationships between these post-course variables and the teacher/course ratings is taken into consideration. Post-course variables are representative of the effects of students' involvement with the course. Consequently, these variables were expected to be strongly related to teacher/course ratings. However, the magnitude of the relationship seemed relatively small in view of the strong zero-order relationships between each of the predictor items and the criterion variables. This must be attributed to the intercorrelations of the predictor variables at this last level of the hierarchical analyses. Even though the normalized regression coefficients used in these hierarchical analyses were not ordered optimally, some information can be gained by considering the absolute magnitude of each beta weight as compared to others at that level. When this comparison is made, it is interesting to note that the "amount learned", "interest" in the course

and often students' perceptions of the "relevance" of course material had the largest beta weights. In fact, the beta weights were consistently several orders of magnitude larger than those observed for student "effort" or difficulty of "workload". Even with the optimal ordering, it would seem very unlikely that either "effort" or "workload" would contribute anything to the variance of the criterion variables. Furthermore, students' "interest" in the course material seemed of major importance in the predictor equations. In effect, rated teacher "skill", "rapport" with students, "content" coverage", and "organization/structure" all became more positively rated as students' impression of the "amount learned", "interest" in the course material, and the "relevance" of course material increased. These effects were most apparent when predicting the composite "content" rating variable.

Pre-course motivation-related variables on the other hand, were useful in predicting only the composite course "content" variable. In these instances, the proportion of variance accounted for was approximately twenty percent or less. Furthermore, the beta weights for the predictor variables were much more likely to be negative and to vary in magnitude from equation to equation. It was more difficult to determine whether or not any one variable might

be more meaningful due to its magnitude and consistency in equations from group to group. This might be suggestive of students attempting to make judgements about the course based upon insufficient information and, therefore, depending upon very different subjective criteria from student to student to make their judgements. The exceptional character of these variables in predicting course "content" might be the result of students having taken the time to inquire about the content prior to the course. Most students probably do at least read course descriptions before enrolling in a course.

Linear combinations of background variables did not predict a meaningful or significant proportion of the variance in any of the teacher or course rating variables. The type of variable, general or composite, was of no importance in these analyses. Furthermore, these findings were expected. Considerable literature was reviewed earlier which would lead one to anticipate the lack of predictive power associated with linear combinations of "GPA", "college year", "major", "class size", and "optionality". The results obtained from the zero-order correlations were also indicative of this result.

At best, the results from these regression analyses have clarified to some degree the meaning of the teacher/course evaluation process. Composite scores, or at

least rated "content", can be biased by students' pre-course attitudes, which would seem to render this rating item less desirable than its general "content" item counterpart. Apparently, asking students' opinions about many specific aspects of a particular course is more likely to produce results dependent upon extraneous criteria: these criteria may or may not be course related. What seems apparent is that these criteria do not appear to be under the control of the teaching faculty member, although the evidence discussed here is not severe enough to recommend that general items are better. Indeed, the most important rating item, that of teacher "skill", was not predictable to any extent by the pre-course characteristics. As a matter of fact, the weight of what evidence there is suggests that the general "skill" rating may be biased to a small degree $(\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}\leq.15)$.

The second group of hierarchical regression analyses was based upon substantially larger sample sizes and was, therefore, more powerful. The purpose was to determine if general and composite rating items could be predicted to any extent by students' post-course motivation-related variables and their opinions of the teacher's "enthusiasm" and the "entertainment" quality of his/her lectures. However, in these analyses the focus was on accounting for variance in general (composite) ratings over and above that

already accounted for by composite (general) items. It is true that teacher/course ratings are predictable from these post-course variables, teacher "enthusiasm" and the "entertainment" value of lectures, but not to a great extent.

General teacher/course rating items were highly predictable from linear combinations of the composite scores. Similarly, each composte rating item was highly predictable from linear combinations of the general rating items. Neither group of analyses proved to have a noticeable advantage in predictive power over the other. However, the effect of post-course motivation-related variables, teacher "enthusiasm", and lecture "entertainment" had differential effects upon general and composite rating items. The latter two variables had an appreciable effect when predicting general ratings of "skill" and "rapport", an effect over and above that variance accounted for by the composite items and the post-course motivation-related variables. Neither "enthusiasm" nor "entertainment" had an effect upon the composite scores. The post-course motivation-related characteristics accounted for considerable variance over and above that accounted for by the linear group of general rating items when predicting the composite "content" and to a lesser extent "organization/structure" rating items. The predictive power

of these post-course variables was particularly strong when predicting the course "content" rating (accounting for as much as thirty-two percent of the variance in the criterion variable). Therefore, to some extent the composite ratings of "skill" and "rapport" seem to provide better estimates of the teacher's classroom performance because they are generalizable over the "enthusiasm" and "entertainment" variables.

A final point of considerable interest pertains to the finding that incomplete composite scores, incomplete factor scores, and complete factor scores produced identical results throughout the analyses in this dissertation. This suggests that one need not consider the use of sophisticated analytic techniques, like factor analysis, to generate composite scores. In fact, all one need do is determine which items are common to a particular teacher/course dimension and add them up to obtain a good estimate of the teacher's performance. This is particularly true in the case of teacher "skill" and "rapport" variables. The major problem is first determining what the appropriate items should be. It would seem that this requires considerable effort in writing items and then selecting an appropriate subset of items to represent a teacher dimension. Analytic tools important to this procedure are item and factor analytic tech-

niques. Presumably, the reason why optimal weighting fails is that the regression model (and factor analysis models) tend to over fit the data, due to capitalization upon chance and the presence of "outliers" in the original data samples (Wainer, 1976). "Outliers" are data points which deviate from multivariate normality. Equal weights avoid both of these problems for several reasons: first, the equal weights are not estimated with the data and, therefore, are less likely to capitalize upon chance factor, and second, the existences of "outliers" in the data have no effect upon the estimates (equal weights) and cannot pull them away from the correct values. As Wainer says, "estimating coefficients in linear models: it don't make no never mind."

The implications of this inital section of research results pertaining to the utility of teacher/course ratings are considerable. However, a final discussion of how these data should be interpreted, concerning which types of rating schemes provide the most useful data to decision makers, should be delayed until after the discussion of the results of similar analyses upon data obtained from samples of students in different disciplines. It is to the second group of analyses that the focus of this dissertation will now to directed.

SECTION II
GROUPED DATA ANALYSES

The purpose of Section II is to determine whether or not the important results observed for the economics course sections also tend to be consistently observed within other disciplines, namely statistics and German. However, several problems in collecting the data prevented any consideration of within-course section analyses for either of these additional disciplines. The primary problem was one of a "natural" reduction in the sample sizes due to student abseteeism on one or both days that the rating data were collected. The average sample sizes for the statistics and German classes prior to the "natural selection process" were 34 and 20, respectively. The reduced sample sizes might have been adequate for estimating bivariate relationships within these course sections, however, one additional problem further reduced the likelihood of any within-sections analyses: many students failed to respond to all of the questions on each questionnaire. Sample sizes within the different classes were as small as seven subjects. Because of these problems, the decision was made to pool student rating data over classes and teachers within each discipline. Thus, the analyses of this section are those presented in Section I (within economics class section analyses) but performed upon the pooled data from each discipline

(ecomonics, statistics, and German). Comparisons, then will be made between groups in the following manner.

Results from the pooled analyses for the economics group will be compared with the results observed in Section I for the separate economics classes. Assuming that the comparisons show that the grouped data has the same overall pattern of results as observed for classes in Section I, then the grouped data results for economics can be compared to results obtained for both statistics and German. However, a few words of caution concerning the use of pooled data analyses: merging the data in the manner described will cause difficulty in interpreting results since it confounds the effect of different instructors within the discipline. In fact, this type of pooled data may be more a cause for confusion than for explanation of relationships between variables. More will be said concerning this issue in the discussion section.

Method

Subjects and Administration. The subject, the administration of questionnaires, the composition of the instruments themselves, and the specification of the hypotheses (listing the appropriate analyses) have already been described in Section I. To reiterate an important point, the only change in the section from the procedures in Section I is that the data for each of the three disciplines has been pooled--

collapsed over teachers and class sections -- rather than using the more desirable within-section analysis. Results observed to be consistent over each of the four economics classes (refer to the results summary in Section I, p.46). will be examined to determine whether or not they replicate for statistics and German groups. Because the procedures are identical, they will not be restated here. The reader can refer back to the appropriate headings in Section I for specific details.

Results

The zero-order correlations between composite and factor-score items with identical meaning once again approached unity as indicated by the range of correlation coefficients ($.89 \leq r \leq .99$). Therefore, the discussion in this section will focus only upon results obtained for equally-weighted "incomplete" composite scores (ICS). Little or no information would be gained by presenting data for the factor-score variables since the zero-order correlations are so large.

The correlations between general items and composite scores with identical meanings were significant in every instance ($.37 \leq r \leq .82$). The smaller correlation coefficients were observed between "organization/structure" variables (GX with X1) for the economics and German groups, .37 and .43, respectively. All other correlation coefficients

exceeded .50 (refer to Table 1-12A and B). The results of the correlational analyses indicate, as in Section I, that "skill" and "rapport" are the two teacher dimensions which are most closely related between general and composite ratings: $r_{GS,\ S1} = .75$ and $r_{GR,\ R1} = .64$. The correlations observed for both "content" ($r_{GC,\ C1} = .55$) and "organization/structure" ($r_{GX,\ X1} = .48$) were less substantial. In the latter instance, this was probably due to the inadequate representation of items inquiring about the "organization/structure" features (exam types, papers required, etc.). However, the correlations were substantial enough to conclude that the general and composite items were measuring similar teacher/course characteristics.

Zero-order correlations and general rating items. Results in this section were considered meaningful when the correlation coefficient was approximately .30 or larger. In some instances, to make explanation clearer, smaller coefficients (when significant) were mentioned. There are two reasons for this: first, the proportion of variance accounted for by smaller zero-order correlation coefficients is relatively uninteresting; and second, the large sample size for the pooled economics group ($N_{pre} = 330$ and $N_{post} = 540$) means that correlations as small as .12 would be significant beyond the $p \leq .01$ level, even though not very meaningful from a variance accounted for

perspective, $r^2 = .01$. It is, perhaps, not entirely coincidental that the size of the coefficient selected for discussion is the same size coefficient found to be significant beyond $p \leq .01$ in both the statistics and German data.

Background variables ("GPA", "college year", "major", "optionality", and "class size") did not typically account for a meaningful proportion of variance in the general rating item or the correlation was not significant. This result was consistent within the three groups ($r^2 \leq .04$). The results reported here for student background variables and general rating items are consistent with results reported for the same variables in Section I. In addition, pre-course motivation-related variables ("amount expected to learn", "relevance", "workload", "interest", and effort") were not strongly related to general ratings in the economics group ($r^2 \leq .06$), nor were they significantly related to general ratings in the German group ($r^2 \leq .10$). However, several pre-course motivation-related variables did correlate highly and significantly with general ratings in the statistics group ($.35 \leq r \leq .49$). "Amount expected to learn" correlated significantly ($p \leq .01$) with "skill" (GS), "rapport" (GR), "content" (GC), and "organization/structure" (GX) (.38, .40, .49, and .38, respectively). Perceived "relevance" and "interest" each correlated with rated "content" (GC), .38 and .35, respectively. In fact, the correlations for

each of these latter two pre-course items with the remaining general items (GS, GR, and GX) all approached significance ($.20 \leq r \leq .33$). It appears that pre-course motivation-related variables can affect teacher/course ratings in certain types of courses (e.g., statistics). The results of these pooled analyses are not contrary to those reported for the economics courses in Section I. Alternatively, these results are as hypothesized: general ratings would be strongly related to students' pre-course attitudes.

In Section I, certain post-course variables ("amount learned", "interest", "enthusiasm", and "entertainment") correlated highly with general rating items. Identical results were also obtained within each of the pooled groups. All four post-course items were strongly ($p \leq .01$) related to each general rating ($.31 \leq r \leq .78$). The only two exceptions occurred in German where "interest" correlated less significantly ($p \leq .05$) with rated "skill" (GS), "content" (GC), and "organization/structure" (GX), .29, .29, and .27, respectively. These results are similar to those reported in Section I, but a bit more consistent here.

One last point before going on to relationships with composite items. Students' assessment of their "expected grade" correlated significantly ($p \leq .01$) with each of the general teacher/course ratings in the statistics group (.34, .28, .42, and .39, respectively), again a result not observed in the economics data of Section I.

Zero-order correlations with composite scores. With few exceptions, the results reported above for general items held true for the composite scores. Background and pre-course motivation-related variables were unrelated to composite score ratings of teachers/courses in both the German and economics groups. This latter finding is at variance with those reported for composite items in Section I. The fact is that both "amount expected to learn" and "interest" were significantly ($p \leq .01$) related to composite items for the economics group but the variance accounted for was less than eight percent. However, for the statistics group "amount expected to learn" was significantly related to the composite scores (.40, .38, .47, and .34, respectively). In addition, "relevance" and "interest" were related to rated course "content" (C1); the correlations were .45 and .47, respectively.

Similarly, the post-course ratings maintained the same pattern of results as reported for general items except that course "organization/structure" (X1) did not correlate significantly with "amount learned", teacher "enthusiasm" and "entertainment" value of lectures, as was true for the general rating (GX). Other than these exceptions, the results are very similar although the magnitude of the relationships may be slightly smaller for the composite scores ($.29 \leq r \leq .73$).

Regression analyses and student background, pre-course and post-course motivation-related variables. Analysis of the pooled data in each group produced consistent results between groups as well as consistent with those reported in Section I. This was true for both sets of regression analyses. In the first case, background variables (B), pre-course (P) and post-course (P) motivation-related variables predicted general or composite rating items. In the second case, the regression analyses used composite (general) items and post-course variables to predict general (composite) teacher/course ratings. Each analysis will be discussed in turn.

Background variables accounted for little or no variance in the criterion variables for either general or composite items (refer to Table 1-14A and B). Pre-course motivation-related variables were not related to general items, but accounted for considerable variance in the "content" composite item ($R^2_{stat} = .28$ and $R^2_{econ} = .18$) in both statistics and economics classes. These results are similar to those reported in Section I for the separate economics sections: the magnitude of the relationships are approximately the same ($.16 \leq R^2_{y.PB} - R^2_{y.B} \leq .22$) Post-course motivation-related variables accounted for considerable variance in the general rating items of "skill", "rapport", and "content" ($.17 \leq R^2_{y.PPB} - R^2_{y.PB} \leq .34$) in the statistics, German and economics groups. However, in several cases, particularly

German, the small sample sizes did not provide enough power to reach significance. The same pattern of results was observed for the same relationships in the statistics and economics groups when predicting composite items, although the variance accounted for tended to be slightly larger [$.10 \leq (\tilde{R}^2_{y.\underline{P}PB} - \tilde{R}^2_{y.PB}) \leq .57$, $p \leq .01$]. The variance accounted for by these relationships for the German data were non-significant and small ($.09 \leq \tilde{R}^2_{y.\underline{P}PB} - \tilde{R}^2_{y.PB} \leq .11$) with the exception of post-course variables and rated course "content". In this case, $\tilde{R}^2_{y.\underline{P}PB} - \tilde{R}^2_{y.PB} = .45$, $F(5, 29) = 4.74$, $p \leq .01$. It appears that composite scores are more likely to be influenced by students' pre- and post-course attitudes than are general rating items with similar meaning.

The regression analyses designed to measure the contributions of post-course motivation-related variables, teacher "enthusiasm", and "entertainment" value of lectures, above and beyond that accounted for by composite (general) items also produced results similar to those reported in Section I. Linear combinations of composite (general) items proved to be strongly related to general (composite) items as indicated by the squared semi-partial correlations ($.41 \leq \tilde{R}^2_{y.C} \leq .70$, $.47 \leq \tilde{R}^2_{y.G} \leq .79$, $p \leq .01$). The only exceptions were for the German and economics groups when predicting the "organization/structure" composite score (the squared semi-partial correlations were .10 and .24 respectively,

as shown in Table 1-15A and B). Broadly speaking, the magnitudes of the relationships for the grouped data were much larger than those reported for the separate economics sections. In addition, and as reported in Section I, post-course motivation-related variables were unrelated to general or composite items, with the exception of the "content" composite item. In this case, the post-course motivation-related variables accounted for (adjusted R^2) 26%, 40%, and 24% of the variance in the "content" ratings. Similarly, "enthusiasm" and "entertainment" had their greatest effects upon the general rating of teacher "skill". The adjusted squared semi-partial correlations were .08, .22, and .09 respectively. All are significant beyond the $p \leq .01$ level. The variance accounted for by "enthusiasm" and"entertainment" in the general "rapport" items were 3%, 5%, and 5%, respectively. The remaining observed results accounted for less than 3% of the variance in the criterion variables and, more often than not, accounted for no variance whatsoever.

The last result to be reported here concerns the linear relationship between background and pre-course motivation-related variables. Just as described in Section I, background variables ("GPA", "college year", "major", "optionality", and "class size") were, for the most part, unrelated to the "amount expected to learn", perceived "relevance" of the material, difficulty of the "workload",

"interest" in the material and "effort" expended for the course. Again, these results are identical to those reported in Section I: the reader is referred to Table 1-16.

Discussion

One important distinction between the data of this section and those of Section I is that these last analyses were based upon ratings pooled over teachers and courses within each discipline (economics, statistics, and German). Pooling the data in this manner may have affected the results of the correlation and regression analyses. In effect, the pooling operation maximizes the variance in the observed data, and, as a result, the observed correlations are expected to be larger on the average than those observed in Section I. For this reason, it is perhaps more important in this section to discuss the similarities in the overall pattern of results as compared to those reported to occur within each economics class. However, there are problems with this approach as well. For example, when a correlational relationship is observed to be large and significant in this section, it could be to a large extent the result of side effects produced by pooling the data. The most important criticism of this approach is that, even though the correlation is moderately large and significant, it may not be consistent within each class section. (This is not a problem for the economics

data since the within-class results are available for comparison.) For example, the correlations could be near zero in some classes and very strong in others; the combined effect being a moderately large relationship. Therefore, it cannot be argued that an observed relationship based upon pooled data generalizes to each class with the sample. The best approach to interpreting the results when dealing with pooled data may be to point out the possibility of consistent relationships between variables within each class section while being sure to emphasize the need for additional data collection or analyses, or both. Therefore, the discussion of results in this portion of the thesis are to be viewed by the reader as being suggestive of important relationships which may be consistent from class to class and discipline to discipline. However, more research is needed to varify the empirical findings observed in the pooled data reported in this section. It is important to keep these words of caution in mind when reading the remainder of this discussion.

Zero-order correlations. The correlational analysis of the pooled economics data were entirely consistent with those reported for the within-class analyses of Section I. In particular, student background variables were not meaningfully related to either general or composite teacher/course ratings, and student's pre-course ratings of "amount expected to learn" and "interest" were again significantly correlated with composite items. However, each of these pre-course variables

were also significantly correlated with general rating items. Only sporadic relationships of this type were observed in Section I within the economics classes. Therefore, the significant observations recorded between pre-course motivation-related variables and general ratings are probably spurious, due to the maximization of variance in the pooled sample. However, this maximization occurs predominantly because pooling the data increases the range of the rating responses on each of the rating scales. The correlations were not very large, typically accounting for about four percent of the variance in the teacher/course ratings -- an increase in line with what might be expected by pooling data, as pointed out earlier. Additional evidence to support this finding is apparent in the relationship between the post-course motivation-related variables and teacher/course ratings.

In the pooled data the correlations are considerably larger. For example, in the results of Section I students' ratings of the "amount learned", "interest", "enthusiasm", and "entertainment" were each highly correlated with both general and specific ratings within classes. The largest within-class relationship accounted for fourty-eight percent of the variance between variables ($r = .69$). While the same pattern of relationships was observed for the pooled data of Section II, the largest (the same relationship as reported above) accounted for 52% of the variance between the two variables ($r = .72$). This is an increase in variance ac-

counted for of four percent over the largest relationship between the same variables in Section I. This increase must be attributed to the effects of pooling the data across classes. Similarly, students' post-course assessments of the "amount learned", "interest", "enthusiasm", and "entertainment" were significantly correlated with each of the general and composite teacher ratings: "skill", "rapport", "content", and "organization/structure", respectively. While these results are consistent with those reported in Section I, they are again larger in magnitude; also undoubtedly a result of pooling the data.

The consistency between the pooled and the within-class section (economics) results was expected since the within-class data proved to be very consistent from class section to class section. Clearly, the results obtained for the pooled economics data can be used as a "standard" of comparison for the results obtained in other sections. Of particular interest were the similarities and differences observed between the economics, statistics, and German data.

Students' pre-course ratings of the "amount expected to learn" was much more strongly related to both general and composite ratings in the statistics group ($r^2 \leq .25$). The same effects were observed for the economics group, although smaller in magnitude than those reported for statistics. Similarly, perceived "relevance" and "interest" in the course were typically significantly related to general and composite

ratings, but the relationships were not quite as large in magnitude ($r^2 \leq .22$). The point here is that the strength of the relationships was large enough to make them unlikely as a result of pooling the data. General and composite ratings were not related to students' pre-course attitudes.

To a limited extent, these findings suggest that there may be differential patterns of results observed in teacher/course ratings, based on the discipline of study. The patterns of correlational results were not the same in all three disciplines. Alternatively, the differing results between German and the other disciplines might be a consequence of the type of teacher who is attracted to the study of languages rather than the social sciences. Such differences could manifest themselves in any number of variables, like teaching style and/or personality characteristics. One other possible reason for the observed disparities between groups might be the composition of the rater populations. Even though the background variables used in this study had no meaningful effect upon pre- and post-course attitudes or teacher/course ratings, it is possible that other descriptive population criteria might be responsible for the differences in observed ratings. Some indication that this might be true is that most of the students in the statistics and economics courses were Business School majors while those in German majored in some area of the College of Arts and Sciences. In any event, these questions cannot be

resolved from the current data and must await additional research.

Both general and composite ratings can apparently be sampled without regard for students' assessments of the desirability of the "workload" in a course or the "effort" the student is willing to invest in the course. These results also seem to be generalizable over disciplines (economics, statistics, and German).

As far as post-course motivation-related variables are concerned, the results observed are very consistent over each discipline and coincide nicely with findings in Section I. In particular, students' ratings of the "amount learned", "interest" in the material, teacher "enthusiasm", and lecture "entertainment" were highly related to both general and composite teacher/course ratings: teacher/course ratings are not generalizable over these items. However, the same argument can be brought to bear on these findings as those used in Section I. These post-course motivation-related variables reflect qualities of the educational experience which should be related to the end-of-semester ratings.

Students' ratings of "expected grade" were related to both general and composite ratings for the statistics group only. This finding can be viewed in a positive or negative light. On the positive side, one would like to think that those students who learned more would perceive the course

as being more favorably taught. Indeed, many researchers have set out to prove just that (Grasha 1972; Kulik and Kulik, 1974; Kulik and McKeachie, 1973; to mention a few). However, it might be that the correlation between "expected grade" and teacher/course ratings reflects something other than student achievement. For example, it might be that grading procedures were biased in the students' favor (e.g., students expected to receive better grades than deserved), and students reciprocated by giving more positive teacher ratings, as Rodin and Rodin (1972) have argued. Unfortunately, achievement measures could not be obtained for students in this study, and so the meaning of the correlations between expected grade and teacher/course ratings cannot be determined from the current data.

Regression analyses. The regression analyses for the grouped data revealed patterns of results very consistent with those reported for the within-class analyses of Section I. Even though the patterns of results were very similar, the small sample sizes in the statistics and German groups (N_S = 51, N_G = 45, respectively) rendered the first set of regression analyses virtually powerless. The first set of analyses used background, pre- and post-course variables as predictors of general and composite ratings. In many instances, large semi-partial correlations were non-significant. Furthermore, the small sample sizes to a great extent insured

that these large squared semi-partial correlations would be adjusted (shrunken R^2) to near zero. One additional point to mention is the inflated proportion of variance accounted for in the economics data; presumably due to the effects of pooling the data within disciplines. The results observed for both statistics and German are probably also inflated.

Linear combinations of background variables ("GPA", "college year", "major", "class size", and "optionality") did not predict a meaningful, significant proportion of variance in any of the general or composite criterion variables. Apparently, background variables are not related to teacher/course ratings, regardless of the course discipline. The variance added by linear combinations of pre-course motivation-related variables ("amount expected to learn", "relevance", "workload", "interest", and "effort") was nominal except for course "content". In this instance, the proportion of variance accounted for was largest in the statistics group ($r = .33$) but not significant. Even though the effect seems consistent over disciplines, a definite conclusion concerning the consistency within course sections as well as the cause of the relationships will require additional research.

Post-course motivation-related variables produced the biggest effects with both general and composite rating items. Apparently, students' assessments of the "amount learned" in actuality, their "interest" in the course material, and to a lesser extent, the "relevance" of the material can influence

the students' assessments of the teacher's ability or "skill", his/her "rapport" with students, the "content" coverage, and the "organization/structure" of the course. The "relevance" variable seems more important (larger in magnitude) in the relationships for the statistics and the German groups, in that order. (This statement is justified since the variance in corresponding variables between groups seemed similar in magnitude if not identical, refer to Appendix B.) The difficulty of the "workload" and the "effort" students expended seemed to be of lesser importance, although the hierarchical analysis does not permit a definite conclusion as to the importance of each of the variables. However, the relative unimportance of these items from class to class, discipline to discipline seems obvious from these data.

In the second set of regression analyses, the pair of variables which again accounted for considerable variance in general rating of teacher "skill" and "rapport" were teacher "enthusiasm" and lecture "entertainment". This was particularly true in the German group data where the two variables accounted for twenty-two percent of the variance in the "skill" rating over and above that predicted by composite and post-course motivation-related characteristics. However, whether this was due to some facet of the discipline or to the type of teacher the discipline attracts cannot be determined from this data.

Also consistent over disciplines was the strong pre-

dictive power of the post-course motivation-related variables with rated course "content" in the three disciplines. The stronger effects appeared when predicting the "content" rating for the German group. In addition, and as noted in the previous discussion, general ratings of course "content" do not seem to be dependent upon students' assessments of "amount learned", "relevance", or "interest". Just what students are rating in this instance is not totally clear. However, the rated "content" coverage did correlate strongly with general teacher ability ("skill"), as did each of the other general ratings. It might be that students base a considerable portion of their ratings of "content" upon the teacher's "skill". When the teacher is perceived as being very capable, ratings of course "content" are higher, that is, more favorable. It could very well be that better teachers are more skilled at selecting and presenting course material. Even though course "content" coverage was designed to be equal from class to class within each discipline, individual teachers still had the option of picking the material used to explain particular concepts. One cannot be sure that the content was actually identical from class to class within the disciplines.

General Discussion and Conclusions

This research has produced several interesting findings which have certain implications for the use of students'

ratings of college teachers and courses. First, the data suggest that background variables, at least those investigated here, are probably not strongly related to teacher/course ratings. Therefore, faculty members need not be concerned that these student typologies (over which they have no control) could produce undesirable affects upon the students' assessments of their abilities. The "optionality" issue, however, is still an open one deserving of additional research. It has been suggested earlier that rating forms contain a question simply asking students, "How much do you want to take this course?" Data obtained for such a question may produce much stronger relationships with end-of-semester teacher/course ratings. For now, the important point is that both general and specific teacher/course ratings are generalizable over background ratings.

Pre-course motivation-related variables were observed to be somewhat related to composite ratings of course "content". However, this is not necessarily a source of contamination since it may reflect students' awareness prior to enrollment of the material to be covered in the course as well as their expectations about how they will perform during the course. In addition, the extent of the relationship between pre-course motivation-related variables with end-of-semester teacher/course ratings may be substantially determined by the type of course material or discipline (e.g., results obtained for statistics

classes). The effects in one instance were strongly associated with ratings of general teacher ability or "skill", which could be undesirable. However, the data in this case were obtained from a pooled sample which can produce misleading results. Firm conclusions concerning the effects of the discipline itself upon such pre-course items must await additional research.

A further assessment by students of the same motivation-related variables at the end-of-semester (post-course variables) produced impressive results. Particularly conclusive were the combined effects of these variables upon rated teacher "skill", whether a general rating item or a composite of specific items. These results are similar to those reported by both Doyle and Whitely (1974) and Granzin and Painter (1973); however, a couple of things need to be mentioned about the latter study. First, these authors used difference scores in an analysis to infer change in student attitude during the course of the semester. In fact, Granzin and Painter argued that their data showed a positive change in student attitude which is reflected in the teacher/course ratings. In the current study, students' attitudes were also found to have changed in that the students became more homogeneous in their feelings about the course they participated in. However, the results were not positive, as indicated by mean ratings (refer to Appendix B). In every case, the ratings

were lower, or less positive, by semester's end. Perhaps this reflects the fact that some students' pre-course attitudes were confirmed while others began the course with no set opinions in mind. In any event, these results are at variance with those of Granzin and Painter, and their results are probably in error due to their measurement method (raw difference scores). The second point to be mentioned is that this change of behavior indicates that teachers do have some control over students' attitudes. One can argue that, depending on the teacher, the ratings would become more positive as a result of carrying through the semester's course.

Can either general or specific ratings be singled out as "better" that the other, based upon the evidence here? Probably not. The reasons for this indefiniteness are many, some of which have been mentioned above. On the basis of the effects of background, pre-course and post-course motivation-related variables upon teacher/course ratings, one must argue that general and specific ratings are about the same in the desirability of their properties with a few qualifications. The major difference is that general items may be a bit more sensitive to teacher "enthusiasm" and "entertainment". However, it remains an open question as to whether this sensitivity should be considered desirable. It may very well be that entertaining teachers stimulate students to work harder and, therefore,

to learn more. If this were true, then the "entertainment" value of the course would be important and desirable. However, if it were true, as Naftulin, , Ware, and Donnelly (1973) suggest, that highly entertaining teachers can "seduce" students into believing that they have learned ("an illusion of having learned", p.74) when little substance has actually been communicated, then the effects of teacher "entertainment" should be perceived in a less positive light. Their belief is that the direct result of a high "entertainment" level in lecture presentations, regardless of the amount of substantial material, is highly favorable teacher ratings. The degree of truth to these assertions is open to debate and deserving of additional empirical investigation. Neither Naftulin, et al., nor the follow-up study by Ware and Williams (1975) provide appropriate data to verify their claims about the effects of teacher entertainment and student learning or students' ratings of teacher performance. Unfortunately, the current data have also failed to provide evidence that would either discredit or verify Naftulin's, et al. claims. Therefore, the case against general ratings has some basis in fact, but insufficient criticism has been marshalled to recommend discontinuing the use of general rating items based upon their strong relationship to teacher "entertainment".

The major point in support of the continued use of general ratings is the simplicity with which they are obtained, but a cautious use is urged since students do not seem to

have a clear conception of what "good teaching" means. However, the evidence observed concerning specific items suggests that a concise understanding of what characteristics define a "good teacher/course", and more importantly, what criteria should be used to assess those qualities does not, as of yet, exist. Perhaps several additional comments will clarify this statement.

The specific items in this research were selected from published literature because they described the commonly accepted dimensions of college teaching which are: teacher "skill"; "rapport"; "content"; and "organization/structure". However, the items did not seem to combine factorially to produce consistent results from class to class (more will be mentioned on this topic in the Addendum on maximum likelihood factor analysis). Of course, the problem could be one of having selected the wrong specific items to measure the dimension of instruction/course intended. If this is the case, then this researcher and many others in the educational psychology profession do not know what characteristics represent quality in teachers/courses; not necessarily an unlikely hypothesis. The other alternative is that the students do not know or do not have a common conception of what constitutes quality teaching which caused their rating data to produce variable results. This author believes the latter to be the case, but realizes the chances that both alternatives are partially true: neither the student nor the

faculty member can know precisely what qualities are essential to a "good teacher" from course to course and discipline to discipline. The bulk of the evidence reported in the literature review and elsewhere (Bettencourt, 1974), however, suggests that students may be slightly more adept at recognizing good classrom teaching performance.

All of this points to a major flaw inherent in the use of specific items in the decision process for tenuring faculty. The problem of which items to use in assessing quality teaching becomes quite complex when one thinks about all the different types of courses and learning situations that require different teaching methods. The problem is sufficiently complex to make one realize that all that can be hoped for is some set of items that approximately describes "good instruction". Assuming that this set of items is available, then composite rating items have some benefits. First of all, evidence cited as a result of this research implies that one need not be concerned with developing complex optimal weighing formulae to combine scores for sets of items. All one needs to know is which items are conceptually similar and then remember how to add. This simplifies the interpretation of data obtained from instruments composed of specific rating items; factor scores are not necessary. Secondly, specific items require students to make a decision about multiple facets of a teacher's performance. The evidence reported in this dissertation suggests that students make their evaluation of these facets based upon

obviously important criteria. The "relevance" of the course to their futures and their "interest" in the course material seem to be commonly considered attributes of "good instruction/courses", as well as the more traditional criterion of the "amount learned" by the students.

The third important feature of specific items is that they provide faculty members with well-defined feedback. If a faculty member has certain strengths or weaknesses in the classroom, it is more likely to be diagnosed using specific rating items. In addition, the fact that specific items are less likely to be related to personality variables, such as "entertainment", will probably be viewed more favorably by faculty members. In sum, the choice of which type of questions to use, general or specific, must be made by the members of the academic community collectively. If the data from student ratings are to be used only as "feedback" to the teachers, an instrument composed of specific items would seem more appropriate. However, if rating data are to be used in promotion and tenure decisions, and they should be, then the choice of instrument becomes more a matter of taste. Either type would actually be adequate, although this author believes that instruments composed of specific items have more all-around utility.

The current research has, at best, only begun to explain important relationships between students' attitudes and teacher ratings. Additional research is needed, as

mentioned throughout this dissertation, to further clarify certain issues. For example, it would be useful to try to determine the casual relationships of students' attitudes about the "amount learned" by them, course "relevance", their "interest" in the material, teacher "enthusiasm", and lecture "entertainment" with rated teacher "skill". Furthermore, the validity and, therefore, the usefulness of students' ratings of college teachers would be more clearly understood if a series of path analyses were performed. In addition, further data should be collected to determine the generalizability of ratings within other disciplines, being particularly careful to select course sections to insure within-class comparisons. These comparisons can only be performed if sections have large student enrollments (e.g., $N \geq 200$.)

The present research is particularly useful because it allows one to make informed decisions about the importance of certain items to future research efforts. But the point that becomes immediately apparent from looking over the grand scheme of this research and the analyses presented in this dissertation is that the scope of the project is far too extensive. It seems as if the introduction of each new variable caused an exponential increase in the number of multivariate analyses necessary to obtain any meaningful result. However, in applied research such as this, it is difficult to imagine a study not having multiple independent

and dependent variables. The moral contained in these statements, gained from experience, is that:

One should never tackle vast projects with half vast ideas!

Table 1-1A: Intercorrelation Matrices for Background, Pre-Course Motivation-Related and General Teacher/Course Rating Items Sections E1 and E2

		(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(18)	(22)	(23)	(8)	(9)	(10)	(11)	(12)	(13)	(45)
	GPA	(2)[1]		.84	.04	-.17	-.08	.21	.05	.13	-.04	.12	.13	.13	.09	.09	.16	.06	-.05	.03	-.03	.05	.06	.09	-.04
	GPA Last Semest	(3)	.82		.04	-.14	-.03	.24	.07	.18	-.01	.17	.16	.12	.06	.02	.19	.09	-.02	.03	-.05	.06	.08	.08	-.05
	College Year	(5)	-.01	-.05		.44	.34	.28	.12	.12	-.05	.05	.03	.22	-.01	.05	.27	.25	.13	.09	.02	-.04	-.01	.00	-.13
	Major	(6)	-.10	-.13	.00		.64	.05	.08	.22	.02	-.03	-.10	.15	.01	-.02	.32	.31	.11	-.07	-.02	-.11	-.01	.00	-.07
	Optionality	(7)	-.22	-.12	-.19	.44		.10	.06	.15	-.02	-.09	-.18	.04	-.04	-.09	.49	.38	.04	-.09	-.07	-.16	-.22	-.19	-.15
	Grade Satisfact	(8)	.17	.21	-.12	-.09	-.02		.17	.20	.08	.11	.25	.26	.02	.16	.20	.14	-.09	-.04	-.08	-.09	.01	.04	-.09
	Amount to Learn	(9)	-.02	-.04	.05	.04	.10	.07		.24	-.06	.22	.46	.28	-.08	.14	-.07	.01	-.03	-.04	.08	.09	.16	.13	.18
	Relevance	(10)	.07	.03	-.05	.26	.23	.11	.33		.05	.18	.40	.27	.01	.28	.14	.20	-.06	-.04	-.11	-.04	.01	-.06	.00
	Learn Environmt	(11)	-.09	-.07	-.06	-.08	-.02	.08	.08	-.03		.07	.10	.05	-.17	-.01	-.03	.11	-.07	.02	-.03	-.09	-.02	.04	-.09
	Workload	(12)	.10	.05	.00	.20	-.04	.07	.31	.26	.05		.29	.30	-.07	.11	-.00	.00	-.07	.13	.05	.11	.04	.10	.12
	Interest	(13)	.18	.17	.10	.06	-.07	.26	.37	.41	.01	.37		.28	-.10	.34	.11	-.07	-.03	.00	.00	.08	.10	.13	.29
	Effort	(14)	.02	-.12	.19	.27	.11	.06	.31	.33	-.05	.36	.37		-.08	.32	-.02	.10	.03	.08	-.01	-.03	.13	.18	.13
	Class Size	(15)	-.04	-.11	-.01	.03	.00	.13	-.04	-.05	-.36	.03	.01	.03		.02	-.01	-.04	.07	-.03	.20	.17	.12	.09	-.04
	Like Soc/Sci	(18)	.13	.11	.02	.03	.16	.17	.16	.22	-.04	.32	.41	.13	.01		.02	.07	-.07	-.09	-.06	.01	-.09	-.03	.03
	Average Grade	(22)	.14	.13	-.17	.17	.07	.17	-.09	.07	.00	.14	.21	-.04	-.06	.25		.79	.08	.04	.01	-.05	-.05	-.01	-.12
	Overall T/C Rat	(23)	.15	.13	-.16	.23	.01	.10	-.11	.03	-.01	.23	.23	.03	-.04	.26	.80		.07	-.03	.00	-.05	-.05	.00	-.04
	Enthusiasm	(8)[2]	.04	.16	.03	-.09	-.18	.06	.14	-.08	-.07	.09	.03	-.07	.18	.03	.16	-.09		.42	.61	.49	.50	.47	.23
	Entertainment	(9)	.09	.04	.13	-.05	-.17	-.12	.01	-.12	-.18	-.06	-.07	-.01	.22	-.28	-.21	-.20	.40		.51	.38	.39	.29	.24
	Skill	(10)	.08	.01	.03	.05	-.08	-.09	.25	-.15	-.05	.08	.02	-.06	.23	-.05	.03	-.02	.42	.46		.59	.53	.49	.17
	Rapport	(11)	.15	.17	.10	-.05	-.19	-.20	-.14	-.23	-.06	.13	.08	.00	.27	-.06	-.09	-.03	.52	.46	.59		.44	.47	.20
	Content	(12)	.17	.17	.12	-.06	-.24	.04	.10	-.16	-.20	.13	.11	.01	.03	-.03	-.08	-.01	.39	.37	.47	.37		.75	.31
	Organization/St	(13)	.01	.02	.08	-.04	-.20	.10	.31	-.12	-.05	.15	.26	-.09	.10	.00	-.03	-.03	.43	.39	.51	.42	.62		.34
	Motivation	(45)	.18	.22	-.02	-.12	-.13	.11	-.05	.06	.00	.00	.15	.07	.12	-.12	-.02	-.01	.13	.36	.16	.24	.18	.19	

Pre-Course Ratings

Post-Course Ratings

1---Item numbers correspond to the item numbers on Questionnaire #1:
numbers 2-7+15 are background characteristics,

2---Item numbers correspond to the item numbers on Questionnaire #2:
numbers 2-9 are post-course motivation-related items,
numbers 10-13 are general teacher/course rating items.

Numbers above diagonal are for class E1, below diagonal are for class E2.
Sample sizes and significance levels are listed in Table 2-3.

Table 1-1B: Intercorrelation Matrices for Background, Pre-Course Motivation-Related and General Teacher/Course Rating Items; Sections E3 and E4.

		(2)[1]	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(18)	(22)	(23)	(8)	(9)	(10)	(11)	(12)	(13)	(45)
Pre-Course Ratings	GPA	(2)	.82	-.15	-.16	-.08	.16	-.06	.09	.04	-.02	.04	-.14	-.01	.17	.05	-.01	-.01	-.06	-.18	-.12	-.06	-.11	.14
	GPA Last Semest	(3)	.79	-.12	-.24	-.04	.14	-.04	.13	.04	.01	.09	-.18	.01	.22	.01	-.05	-.13	-.10	-.25	-.19	-.02	-.09	-.05
	College Year	(5)	.01	.03	.05	-.08	-.04	.06	.06	-.09	.14	.08	.21	-.10	.04	-.06	-.03	.16	.10	.04	-.02	-.02	-.05	-.12
	Major	(6)	-.07	-.05	.21	.48	-.03	.18	.30	.01	.10	.03	.20	.18	-.05	.06	.09	.23	.20	.24	.29	.06	.14	-.07
	Optionality	(7)	-.09	-.10	.21	.63	.02	-.07	.16	-.02	-.09	-.22	.07	.15	-.12	-.02	-.03	-.02	-.20	-.05	.02	-.13	-.05	-.11
	Grade Satisfact	(8)	.30	.26	.09	-.05	-.06	.23	.17	.07	.18	.22	.15	-.02	.17	-.03	.00	-.10	.07	-.12	-.02	-.05	.03	.09
	Amount to Learn	(9)	-.10	-.11	-.01	.06	.07	.15	.34	.00	.40	.51	.45	.00	.22	-.07	-.07	.00	.13	.06	.06	.14	.13	-.06
	Relevance	(10)	-.04	-.10	-.02	.18	.11	.13	.27	.07	.30	.44	.24	.11	.28	-.03	.03	.01	.05	-.10	-.03	-.02	.04	.10
	Learn Environmt	(11)	.04	.07	-.03	.03	.02	.02	-.04	.01	-.05	-.02	.08	-.35	.03	-.05	-.04	-.11	-.04	-.02	-.04	.01	.00	-.07
	Workload	(12)	-.05	-.04	-.03	.16	-.01	.09	.24	.21	.09	.44	.39	.01	.24	-.02	.00	.03	.13	-.00	.00	.08	.06	.04
	Interest	(13)	.01	-.01	-.08	-.09	-.17	.34	.44	.29	-.05	.32	.36	.08	.40	.18	.21	.09	.33	.21	.09	.27	.21	.06
	Effort	(14)	-.16	-.17	.16	.24	.14	.02	.35	.27	.08	.33	.26	-.12	.18	-.06	-.05	-.16	-.05	-.06	.04	.07	.08	.06
	Class Size	(15)	-.05	-.06	-.09	.05	-.12	-.04	-.04	.01	-.20	.00	.01	.00	.06	.03	.00	-.03	.01	.07	-.07	.12	.09	-.07
	Like Soc/Sci	(18)	-.03	.01	-.02	.09	-.11	.18	.20	.35	.08	.21	.41	.27	.05	.22	.22	-.12	-.01	-.12	-.17	.06	.01	.01
	Average Grade	(22)	.23	.33	.04	.24	.23	.14	-.05	.03	.14	.02	.03	-.08	-.12	.11	.83	-.05	-.09	-.05	-.13	-.15	-.10	.00
	Overall T/C Rat	(23)	.04	.15	.11	.17	.17	.12	.00	.04	.05	-.03	.09	.08	-.14	.12	.58	.02	-.09	-.05	-.12	-.15	-.13	-.01
Post-Course Ratings	Enthusiasm	(8)[2]	.00	-.11	.11	-.07	-.10	.22	.00	.22	.02	.16	.42	.13	.10	.17	-.06	.07	.48	.58	.60	.41	.36	-.01
	Entertainment	(9)	.01	-.01	.05	.04	-.07	.12	.04	.15	.07	.08	.19	.13	-.07	.24	.08	.13	.54	.69	.56	.50	.55	-.09
	Skill	(10)	-.10	-.16	.08	.17	.09	.12	.15	.26	.01	.13	.30	.16	.06	.24	-.03	.06	.64	.64	.64	.60	.63	-.07
	Rapport	(11)	-.02	-.09	.10	.27	.20	.07	.13	.25	-.11	.22	.27	.24	.04	.20	.01	.05	.53	.46	.55	.43	.50	-.01
	Content	(12)	.09	-.02	0.22	0.22	-.18	.18	.07	.28	.04	.11	.23	.04	.01	.14	-.09	-.08	.50	.44	.60	.49	.58	.01
	Organization/St	(13)	.05	-.10	-.10	-.14	-.17	.16	.08	.13	-.03	.13	.30	.10	.01	.13	-.04	.09	.49	.46	.51	.40	.66	-.03
	Motivation	(45)	-.01	.04	-.12	-.18	-.18	.07	.05	.03	-.01	.06	.20	.03	.06	.11	-.01	-.03	.32	.21	.32	.42	.39	.34

1---Item numbers correspond to the item numbers on Questionnaire #1:
numbers 2-7+15 are background characteristics.
2---Item numbers correspond to the item numbers on Questionnaire #2:
numbers 2-9 are post-course motivation-related items,
numbers 10-13 are general teacher/course rating items.
Numbers above diagonal are for class E3, below diagonal are for class E4.
Sample sizes and significance levels are listed in Table 2-3.

Table 1-1C: Intercorrelation Matrices for Background, Post-Course Motivation-Related and General Teacher/Course Rating Items; Sections E1 and E2.

		(2)	(3)	(5)	(6)	(7)	(88)	(15)	(18)	(22)	(23)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(45)
GPA	(2)[1]		.84	.04	-.17	-.08	.21	.09	.09	.16	.06	.18	.03	.09	-.03	.10	.11	-.05	.03	-.03	.05	.06	.09	-.04
GPA Last Semest	(3)	.82		.04	-.14	-.03	.24	.06	.02	.19	.09	.32	.06	.14	.01	.15	.18	-.02	.03	-.05	.06	.08	.08	-.05
College Year	(5)	-.01	-.05		.44	.34	.28	-.01	.05	.27	.25	.23	-.12	.15	.17	.02	.00	.13	.09	.02	-.04	-.01	.00	-.13
Major	(6)	-.10	-.13	.00		.64	.05	.01	-.02	.32	.31	.08	-.05	.27	.06	.00	.08	.11	-.07	-.02	-.11	-.01	.00	-.07
Optionality	(7)	-.22	-.12	-.19	.44		.10	-.04	-.09	.49	.38	.14	-.06	.23	.10	-.07	.01	.04	-.09	-.07	-.16	-.22	-.19	-.15
Crade Satisfact	(8)	.17	.21	-.12	-.09	-.02		.02	.16	.20	.14	.65	-.06	.15	.04	.14	.07	-.09	-.04	-.08	-.09	.01	.04	-.09
Class Size	(15)	-.04	-.11	-.01	.03	.00	.13		.02	-.01	-.04	.08	-.01	.02	-.09	.10	.08	.07	-.03	.20	.17	.12	.09	-.04
Like Soc/Sci	(18)	.13	.11	.02	.03	-.16	.17	.01		.02	.07	.15	-.09	.14	.02	.12	.08	-.07	-.09	-.06	.01	-.09	-.03	.03
Average Grade	(22)	.14	.13	-.17	.17	.07	.17	-.06	.25		.79	.27	-.12	.24	.03	.11	.03	.08	.04	.01	-.05	-.05	-.01	-.12
Overall T/C Rat	(23)	.15	.13	-.16	.23	.01	.10	-.04	.26	.80		.22	-.09	.13	.03	.12	.00	.07	-.03	.00	-.05	-.05	.00	-.04
Grade Expected	(2)[2]	-.02	.18	-.01	-.07	-.10	.19	-.24	-.12	-.05	-.01		-.01	.22	-.10	.30	.02	-.03	.05	.00	.00	.04	.09	.08
Amount to Learn	(3)	.19	.14	-.03	-.05	-.20	-.02	.23	-.08	-.02	-.05	.14		.12	-.01	.32	.23	.18	.17	.33	.15	.46	.41	.16
Relevance	(4)	.09	.12	-.03	.14	.16	.11	.01	.12	.06	.06	.26	.07		.04	.37	.24	.10	.07	.03	.11	.11	.13	.17
Workload	(5)	.36	.29	.08	-.15	-.21	.31	.13	.28	.03	.03	-.12	.23	.00		-.16	.41	-.02	-.04	-.01	-.11	-.06	-.04	-.22
Interest	(6)	.33	.36	.07	-.15	-.32	.21	.16	.30	.07	.16	.29	.38	.43	.10		.11	.32	.31	.30	.29	.36	.37	.48
Effort	(7)	.36	.28	.16	-.07	-.13	.21	.04	.05	.10	.04	.01	.30	.17	.53	.31		-.12	-.07	-.07	-.04	-.02	.04	.01
Enthusiasm	(8)	.04	.16	.03	-.09	-.18	.06	.18	.03	-.16	-.09	.12	.31	.14	.08	.31	.14		.42	.61	.49	.50	.47	.23
Entertainment	(9)	.09	.04	.13	-.05	-.17	-.12	.22	-.28	-.21	-.20	.04	.23	.14	-.17	.32	.17	.40		.51	.38	.39	.29	.24
Skill	(10)	.08	.01	.03	.05	-.08	-.09	.23	-.05	.03	-.02	-.06	.40	.01	.06	.33	.31	.42	.46		.59	.53	.49	.17
Rapport	(11)	.15	.17	.10	-.05	-.19	-.20	.27	-.06	-.09	-.03	.06	.25	.03	.15	.31	.28	.52	.46	.59		.44	.47	.20
Content	(12)	.17	.17	.12	-.06	-.24	.04	.03	-.03	-.08	-.01	.07	.29	.11	.00	.31	.19	.39	.37	.47	.37		.75	.31
Organization/St	(13)	.01	.02	.08	-.04	-.20	.10	.10	.00	-.03	-.03	.23	.21	.20	-.07	.34	.13	.43	.39	.51	.42	.62		.34
Motivation	(45)	.18	.22	-.02	-.12	-.13	.11	.12	-.12	-.02	-.01	.10	.08	.15	.14	.31	.12	.13	.36	.16	.24	.18	.19	

Pre-Course Ratings

Post-Course Ratings

1---Item numbers correspond to the item numbers on Questionnaire #1: numbers 2-7+15 are background characteristics.

2---Item numbers correspond to the item numbers on Questionnaire #2: numbers 2-9 are post-course motivation-related items, numbers 10-13 are general teacher/course rating items.

Numbers above diagonal are for class E1, below diagonal are for class E2.
Sample sizes and significance levels are listed in Table 2-3.

Table 1-1D: Intercorrelation Matrices for Background, Post-Course Motivation-Related and General Teacher/Course Rating Items; Sections E3 and E4.

		(2)	(3)	(5)	(6)	(7)	(8)	(15)	(18)	(22)	(23)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(45)
Pre-Course Ratings	GPA	(2)[1]		.82	-.15	-.16	-.08	.16	-.01	.17	.05	-.01	.08	-.22	.10	.05	.03	.10	-.01	-.06	-.18	-.12	-.06	-.11	.14
	GPA Last Semest	(3)	.79		-.12	-.24	-.04	.14	.01	.22	.01	-.05	.25	-.24	.12	-.08	-.03	-.06	-.13	-.10	-.25	-.19	-.02	-.09	-.05
	College Year	(5)	.01	.03		.05	-.08	-.04	-.10	.04	-.06	-.03	-.21	.04	-.02	.12	-.04	.17	.16	.10	.04	-.02	-.02	-.05	-.12
	Major	(6)	-.07	-.05	.21		.48	-.03	.18	-.05	.06	.09	-.03	.16	.28	.31	.12	.27	.23	.20	.24	.29	.06	.14	-.07
	Optionality	(7)	-.09	-.10	.21	.63		.02	.15	-.12	-.02	-.03	-.05	-.06	-.01	-.07	-.04	.02	-.02	-.20	-.05	.02	-.13	-.05	-.11
	Grade Satisfact	(8)	.30	.26	.09	-.05	-.06		-.02	.17	-.03	.00	.25	-.07	.07	-.15	-.07	-.07	-.10	.07	-.12	-.02	-.05	.03	.09
	Class Size	(15)	-.05	-.06	-.09	.05	-.12	-.04		.06	.03	.00	.05	.00	.04	.00	.02	-.09	-.03	.01	.07	-.07	.12	.09	-.07
	Like Soc/Sci	(18)	-.03	.01	-.02	.08	-.11	.18	.05		.22	.22	-.02	-.09	.18	-.05	.05	-.04	-.12	-.01	-.12	-.17	.06	.01	.01
	Average Grade	(22)	.23	.33	.03	.24	.23	.14	-.12	.11		.83	.09	.09	-.03	-.09	.21	.03	-.05	-.09	-.05	-.13	-.15	-.10	.00
	Overall T/C Rat	(23)	.04	.15	.11	.17	.17	.12	-.14	.12	.58		.11	.04	-.01	-.06	.09	.04	.02	-.09	-.07	-.10	-.15	-.13	-.01
Post-Course Ratings	Grade Expected	(2)[2]	.36	.27	.01	.08	.11	.60	.08	.22	.12	.13		.26	.25	-.23	.31	.13	.06	.15	.11	.17	.14	.12	.02
	Amount to Learn	(3)	.05	.01	.04	.03	.08	.20	-.08	.07	.02	.04	.23		.29	.12	.53	.23	.26	.42	.53	.07	.45	.45	.07
	Relevance	(4)	.06	.02	-.06	.01	.15	.12	.08	.30	.27	.27	.24	.33		.17	.41	.14	.10	.20	.16	.19	.16	.29	.14
	Workload	(5)	-.09	-.13	.25	.25	.28	.03	.07	.03	-.04	-.05	.00	.21	.19		.17	.41	.04	-.04	-.01	.19	-.06	.05	.12
	Interest	(6)	.09	.01	-.06	.07	.03	.29	-.01	.26	.15	.27	.36	.53	.47	.23		.43	.27	.37	.37	.34	.43	.46	.15
	Effort	(7)	-.07	-.13	.35	.19	.22	.23	.10	.13	-.14	.03	.11	.38	.25	.49	.36		.14	.04	.05	.05	.01	.07	.13
	Enthusiasm	(8)	.00	-.11	.11	-.07	-.10	.22	.10	.17	-.06	.07	.15	.33	.21	.11	.45	.20		.48	.58	.60	.41	.36	-.01
	Entertainment	(9)	.01	-.01	.05	.04	-.07	.12	-.07	.24	.08	.13	.09	.30	.14	.06	.39	.06	.54		.69	.56	.50	.55	-.04
	Skill	(10)	-.10	-.16	.08	.17	.09	.12	.06	.24	-.03	.06	.15	.45	.16	.03	.51	.19	.64	.64		.64	.60	.63	-.07
	Rapport	(11)	-.02	-.09	.10	.21	.20	.07	.04	.20	.01	.05	.06	.39	.16	.02	.37	.18	.53	.46	.55		.43	.50	-.01
	Content	(12)	.09	-.02	-.22	-.22	-.18	.18	.01	.14	-.09	-.08	.07	.54	.25	.04	.55	.12	.50	.44	.60	.49		.58	.01
	Organization/St	(13)	.05	-.10	-.10	-.14	-.17	.16	.01	.13	-.04	.09	.15	.42	.23	.09	.54	.14	.49	.46	.51	.40	.66		-.03
	Motivation	(45)	-.01	.04	-.12	-.18	-.18	.07	.06	.11	-.01	-.03	.03	.32	.11	-.05	.28	.06	.32	.21	.32	.42	.39	.34	

1---Item numbers correspond to the item numbers on Questionnaire #1: numbers 2 -7+15 are background characteristics.

2---Item numbers correspond to the item numbers on Questionnaire #2: numbers 2-9 are post-course motivation-related items, numbers 10-13 are general teacher/course rating items.

Numbers above diagonal are for class E3, below diagonal are for class E4.
Sample sizes and significance levels are listed in Table 2- 3.

Table 1-2: Sample Sizes and Two-tailed Significance Levels for Correlation Coefficients

	Total N	N for Pre by Post Correlations	r for P<.05	r for P<.01	N for Pre by Pre Correlations	r for P<.05	r for P<.01	N for Post by Post Correlations	r for P<.05	r for P<.01
E1	295	100	.195*	.254	197	.138	.181	188	.138	.181
E2	182	74	.224	.292	140	.159	.208	107	.195	.254
E3	216	80	.211	.275	161	.159	.208	122	.178	.232
E4	269	90	.205	.267	203	.138	.181	142	.159	.208

*Tabled values for r at probability levels taken from Statistical Reasoning in Psychology and Education, by E. W. Minium, 1970 (New York: John Wiley and Sons, Inc.) p. 446.

Table 1-5A: Composite and Factor-Score Variables Correlated with Pre-Course and Post-Course Questionnaire Items; Economics class E1

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
GPA	(2)[3]	-.03	.05	.06	.09	.06	.08	.03	-.20	.08	.09	.03	-.20	.10	.10	.10	-.25
GPA Last Semest	(3)	-.05	.06	.08	.08	.09	.12	.06	-.20	.10	.14	.06	-.23	.11	.14	.09	-.27
College Year	(5)	.02	-.04	-.01	.01	.04	.10	.03	.05	.03	.09	.02	.06	.04	.02	.02	.05
Major	(6)	-.02	-.10	-.01	-.01	.00	.06	.02	.04	-.02	.04	.00	.06	-.02	-.01	.03	.06
Optionality	(7)	-.07	-.16	-.22	-.18	-.15	.00	-.13	-.11	-.14	-.01	-.15	-.09	-.08	-.02	-.12	-.09
Class Size	(15)	.20	.17	.12	.09	.13	.04	.05	.14	.08	.05	.05	.14	-.01	.08	.01	.18
Grade Satisfact	(8)	-.07	-.09	-.01	.04	.05	.07	.09	-.13	.02	.07	.09	-.13	.01	.10	.11	-.15
Amount Learned	(9)	.08	.09	.16	.13	.18	.22	.30	.14	.16	.22	.32	.13	.10	.19	.27	.10
Relevance	(10)	-.11	-.04	.01	-.05	-.16	.09	.22	-.04	-.17	.12	.20	-.04	-.22	.13	.24	-.07
Workload	(12)	.05	.11	.04	.10	.08	.19	.23	.06	.10	.20	.24	.02	-.01	.22	.21	.00
Interest	(13)	.00	.08	.10	.13	.08	.32	.45	.15	.08	.31	.46	.12	-.06	.31	.44	.06
Effort	(14)	-.01	-.03	.13	.18	.05	.25	.20	-.03	.10	.26	.19	-.04	.07	.27	.15	-.11
Like Soc/Sci	(18)	-.06	.01	-.09	-.03	-.07	.17	.13	.07	-.07	.18	.12	.06	-.18	.23	.08	.07
Average Grade	(22)	.01	-.05	-.05	-.01	.07	.06	-.03	-.04	.02	.06	-.05	-.03	.02	.01	-.04	-.03
Overall T/C Rat	(23)	.00	-.05	-.05	.00	.07	.08	-.06	-.04	.02	.06	-.07	-.04	.02	.00	-.03	-.04
Expected Grade	(2)	-.01	.00	.04	.09	.10	.11	.12	.06	.09	.11	.12	.04	.06	.10	.11	.02
Amount Learned	(3)	.33	.15	.46	.41	.33	.26	.34	.30	.35	.27	.35	.32	.22	.14	.27	.26
Relevance	(4)	.03	.11	.11	.13	.04	.09	.40	.13	.05	.07	.37	.15	-.09	-.01	.43	.09
Workload	(5)	-.02	-.11	-.06	-.04	-.24	-.11	-.12	-.26	-.19	-.08	-.14	-.27	-.10	-.01	-.10	-.26
Interest	(6)	.30	.29	.36	.37	.40	.33	.61	.35	.39	.31	.62	.33	.18	.20	.57	.27
Effort	(7)	-.07	-.04	-.02	.03	-.09	-.04	.22	-.02	-.08	-.03	.21	-.03	-.14	.01	.24	-.01
Enthusiasm	(8)	.61	.49	.50	.47	.47	.30	.23	.23	.47	.27	.24	.23	.41	.15	.12	.18
Entertainment	(9)	.51	.38	.38	.29	.36	.32	.27	.12	.37	.30	.29	.14	.29	.22	.22	.08
Motivation	(45)	.16	.20	.31	.34	.38	.22	.41	.23	.40	.19	.42	.21	.28	.06	.39	.16

Pre-Course Ratings

Post-Course Ratings

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.
3---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Pre-course ratings: N=100, $|r| \geq .25$, $P < .01$
Post-course ratings: N=188, $|r| \geq .18$, $P < .01$

Table 1-5B: Composite and Factor-Score Variables Correlated with Pre-Course and Post-Course Questionnaire Items; Economics class E2

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
GPA	(2)[3]	.08	.15	.17	.01	.12	.16	.17	.04	.14	.19	.17	.05	.09	.19	.09	.01
GPA Last Semest	(3)	.01	.17	.17	.02	.10	.18	.17	.02	.07	.20	.17	.01	.03	.23	.11	.01
College Year	(5)	.03	.10	.12	.08	.07	.09	.08	-.01	.08	.07	.09	-.04	.15	.04	.06	.00
Major	(6)	.05	-.05	-.06	-.04	-.06	.10	-.05	.00	-.05	.11	-.07	.03	-.09	.13	-.09	-.04
Optionality	(7)	-.08	-.19	-.24	-.20	-.17	.09	-.16	-.09	-.17	.11	-.18	-.08	-.17	.17	-.19	-.06
Class Size	(15)	.23	.27	.03	.10	.13	.14	.15	.07	.15	.13	.15	.06	.18	.10	.07	.10
Grade Satisfact	(8)	-.09	-.20	.04	.10	-.02	-.17	.12	.05	-.04	-.17	.12	.10	-.06	-.22	.19	.07
Amount Learned	(9)	.25	.14	.11	.32	.16	.01	.06	.06	.16	-.01	.07	.06	.23	-.07	.05	.07
Relevance	(10)	-.15	-.23	-.16	-.12	-.08	-.11	-.05	-.17	-.09	-.08	-.05	-.11	-.03	-.10	.00	-.14
Workload	(12)	.09	.13	.13	.15	.04	-.03	.07	.02	.07	-.03	.08	.02	.08	-.06	.08	-.03
Interest	(13)	.02	.08	.11	.26	.05	-.02	.29	.12	.04	.00	.30	.12	-.03	-.05	.35	.08
Effort	(14)	-.06	.00	.01	-.09	-.15	-.02	.03	.02	-.11	-.02	.03	.04	-.12	-.01	.05	.01
Like Soc/Sci	(18)	-.05	-.06	-.03	.00	-.16	-.11	.13	-.17	-.19	-.08	.14	-.16	-.16	-.05	.20	-.11
Average Grade	(22)	-.04	-.09	-.08	-.03	-.06	-.04	.01	-.07	-.06	-.04	.00	-.07	-.08	-.02	.05	-.10
Overall T/C Rat	(23)	-.02	-.04	-.02	-.03	-.10	-.02	.09	-.08	-.08	-.02	.09	-.09	-.09	-.01	.15	-.12
Expected Grade	(2)	-.06	.06	.07	.23	.09	.08	.13	.07	.06	.09	.13	.08	.01	.03	.14	.01
Amount Learned	(3)	.40	.25	.29	.21	.35	.17	.26	.32	.35	.17	.27	.33	.29	.06	.18	.29
Relevance	(4)	.01	.03	.11	.20	.14	.19	.51	.16	.15	.20	.49	.18	.02	.12	.48	.08
Workload	(5)	.07	.15	.00	-.07	-.08	.17	.00	.11	-.09	.19	.01	.12	-.13	.20	-.04	.11
Interest	(6)	.33	.31	.31	.34	.25	.25	.60	.25	.27	.26	.62	.26	.13	.15	.58	.11
Effort	(7)	.31	.28	.19	.13	.25	.40	.33	.32	.25	.43	.34	.31	.13	.37	.22	.22
Enthusiasm	(8)	.42	.52	.39	.43	.49	.42	.33	.38	.49	.40	.34	.37	.46	.29	.13	.29
Entertainment	(9)	.46	.47	.37	.39	.42	.36	.31	.27	.45	.35	.31	.28	.40	.20	.16	.16
Motivation	(45)	.16	.24	.18	.19	.22	.15	.21	.09	.21	.16	.21	.11	.17	.08	.16	.01

Pre-Course Ratings

Post-Course Ratings

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.
3---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Pre-course ratings: N=74, $|r| \geq .29$, $P < .01$
Post-course ratings: N=107, $|r| \geq .25$, $P < .01$

Table 1-5C: Composite and Factor-Score Variables Correlated with Pre-Course and Post-Course Questionnaire Items; Economics class E3

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
Pre-Course Ratings																	
GPA	(2)[3]	-.18	-.12	-.05	-.11	-.20	-.14	.10	-.06	-.20	-.11	.10	-.06	-.23	-.04	.15	-.03
GPA Last Semest	(3)	-.25	-.19	-.02	-.09	-.20	-.21	.17	.08	-.22	-.20	.15	.09	-.27	-.13	.23	.14
College Year	(5)	.04	-.02	-.02	-.05	.09	.02	.01	-.02	.10	.01	.01	.02	.14	-.07	.00	-.02
Major	(6)	.24	.29	.06	.14	.13	.14	.11	.02	.17	.14	.10	.06	.17	.06	.08	-.04
Optionality	(7)	-.05	.01	-.13	-.05	-.06	-.12	-.17	.04	-.05	-.13	-.15	.02	-.02	-.14	-.09	.05
Class Size	(15)	.07	-.07	.12	.09	.03	-.06	-.02	.04	.06	-.05	-.02	.08	.08	-.12	-.03	.04
Grade Satisfact	(8)	-.12	-.02	-.05	.03	-.11	.02	.08	-.01	-.11	.03	.08	.03	-.15	.08	.10	.03
Amount Learned	(9)	.06	.06	.14	.13	.21	.14	.52	.33	.20	.14	.50	.36	.10	.03	.47	.25
Relevance	(10)	-.10	-.03	-.03	.04	-.02	-.03	.28	.12	.00	-.03	.25	.16	-.03	-.11	.28	.10
Workload	(12)	.00	.00	.08	.06	.03	.20	.26	.08	.04	.17	.24	.13	-.02	.15	.22	.07
Interest	(13)	.21	.10	.27	.21	.23	.18	.43	.33	.21	.17	.43	.38	.11	.08	.38	.29
Effort	(14)	-.06	.04	.07	.08	.01	.01	.21	.13	.02	.00	.19	.19	-.03	.00	.20	.13
Like Soc/Sci	(18)	-.12	-.17	.06	.00	-.10	.02	.23	.13	-.11	.04	.21	.17	-.18	.05	.26	.13
Average Grade	(22)	-.05	-.13	-.15	-.09	-.11	-.09	-.02	.05	-.12	-.11	-.02	-.01	-.14	-.09	.09	.01
Overall T/C Rat	(23)	-.07	-.10	-.15	-.13	-.13	-.05	-.06	-.01	-.14	-.08	-.06	-.04	-.16	-.03	.04	-.01
Post-Course Ratings																	
Expected Grade	(2)	.11	.17	.14	.12	.14	.06	.31	.32	.10	.02	.32	.30	.00	.33	.01	.32
Amount Learned	(3)	.53	.41	.45	.45	.46	.36	.54	.51	.43	.35	.55	.50	.32	.46	.18	.39
Relevance	(4)	.16	.19	.16	.29	.22	.18	.48	.36	.21	.16	.46	.37	.09	.43	.05	.29
Workload	(5)	-.01	.09	-.06	.05	.03	-.04	.12	-.07	.05	-.02	.11	-.06	.05	.11	-.04	-.08
Interest	(6)	.37	.34	.43	.46	.40	.27	.67	.25	.37	.25	.67	.23	.24	.66	.12	.11
Effort	(7)	.05	.05	.01	.07	.08	.05	.30	.03	.09	.05	.30	.06	.04	.31	.04	.01
Enthusiasm	(8)	.58	.60	.41	.36	.51	.48	.22	.18	.48	.47	.24	.22	.47	.05	.31	.11
Entertainment	(9)	.69	.56	.50	.55	.54	.52	.44	.21	.52	.51	.45	.22	.44	.30	.40	.07
Motivation	(45)	-.07	-.01	.01	-.03	-.08	-.05	.21	.08	-.12	-.04	.22	.08	-.18	.24	.01	.10

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.
3---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Pre-course ratings: N=80, $|r| \geq .28$, $P < .01$
Post-course ratings: N=122, $|r| \geq .23$, $P < .01$

Table 1-5D: Composite and Factor-Score Variables Correlated with Pre-Course and Post-Course Questionnaire Items; Economics class E4

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
	GPA (2)[3]	-.10	-.03	.09	.05	.15	.02	-.05	-.08	.16	.01	-.04	-.09	.17	-.03	-.04	-.11
	GPA Last Semest (3)	-.16	-.09	-.02	-.10	.01	-.06	-.05	-.06	.04	-.06	-.05	-.07	.03	-.06	-.04	-.09
	College Year (5)	.07	.10	-.22	-.10	.01	.16	.00	.13	.00	.14	.00	.16	-.01	.13	-.05	.14
	Major (6)	.17	.21	-.22	-.14	-.11	.12	.04	.08	-.11	.16	.04	.07	-.20	.22	.01	.09
	Optionality (7)	.09	.20	-.18	-.17	-.13	.09	.00	.03	-.13	.11	-.01	.04	-.21	.15	-.02	.05
	Class Size (15)	.06	.04	.01	.01	.04	-.02	.08	-.02	.09	.01	.09	-.05	.07	-.02	.10	-.10
	Grade Satisfact (8)	.12	.07	.19	.16	.37	.20	.13	.05	.37	.16	.13	.06	.39	.07	.05	.01
	Amount Learned (9)	.15	.13	.07	.08	.08	.20	.19	.17	.06	.18	.19	.21	-.01	.19	.17	.20
	Relevance (10)	.26	.25	.28	.13	.27	.23	.34	.24	.26	.20	.35	.25	.19	.06	.33	.21
	Workload (12)	.13	.22	.11	.13	.09	.14	.21	.03	.11	.11	.21	.05	.03	.07	.24	.01
	Interest (13)	.30	.27	.23	.30	.31	.29	.43	.35	.31	.25	.44	.37	.20	.11	.41	.32
	Effort (14)	.16	.24	.04	.10	.13	.22	.20	.03	.13	.21	.20	.06	.07	.21	.17	.02
	Like Soc/Sci (18)	.24	.20	.14	.13	.21	.18	.21	.15	.20	.17	.21	.18	.15	.12	.16	.14
	Average Grade (22)	-.03	.01	-.09	-.04	-.08	-.02	.01	.03	-.09	-.03	.00	.03	-.13	-.07	.04	.05
	Overall T/C Rat (23)	.07	.06	-.08	.09	-.05	.08	.11	-.03	-.05	.06	.10	-.04	-.08	.05	.14	-.04
Post-Course Ratings	Expected Grade (2)	.15	.05	.08	.15	.27	.15	.13	.10	.27	.15	.15	.09	.22	.09	.08	.09
	Amount Learned (3)	.45	.39	.54	.42	.46	.36	.55	.55	.43	.34	.56	.53	.28	.17	.47	.51
	Relevance (4)	.16	.16	.25	.23	.19	.14	.56	.18	.18	.14	.55	.16	.01	.05	.60	.12
	Workload (5)	.03	.02	.04	.09	-.02	.14	.21	.04	-.04	.14	.21	.02	-.12	.19	.21	.02
	Interest (6)	.51	.37	.55	.54	.50	.48	.65	.40	.50	.45	.65	.38	.35	.32	.58	.29
	Effort (7)	.70	.18	.12	.13	.09	.18	.29	.11	.11	.19	.29	.08	.03	.14	.29	.04
	Enthusiasm (8)	.64	.53	.50	.49	.63	.48	.49	.44	.62	.44	.50	.42	.53	.27	.34	.37
	Entertainment (9)	.64	.46	.44	.46	.51	.39	.36	.30	.50	.36	.36	.29	.46	.25	.24	.22
	Motivation (45)	.32	.42	.39	.34	.38	.30	.36	.36	.36	.28	.36	.36	.25	.15	.29	.32

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.
3---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Pre-course ratings: N=90, $|r| \geq .28$, $p < .01$
Post-course ratings: N=142, $|r| \geq .22$, $p < .01$

Table 1-6A: Intercorrelation Matrices for General, Composite and Factor Score Ratings; Sections E1 and E2

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
General Items	(GS)		.59	.53	.49	.59	.39	.30	.28	.59	.37	.31	.29	.50	.21	.18	.23
	(GR)	.59		.44	.47	.51	.47	.41	.31	.51	.46	.42	.30	.35	.33	.29	.23
	(GC)	.47	.38		.75	.57	.26	.31	.32	.54	.26	.32	.32	.47	.06	.22	.27
	(GX)	.51	.42	.62		.62	.39	.41	.38	.62	.38	.42	.38	.50	.18	.28	.31
Incomplete Factor Scores	(S1)	.52	.48	.61	.63		.49	.42	.43	.97	.46	.44	.45	.85	.19	.28	.37
	(R1)	.45	.70	.36	.28	.47		.47	.34	.50	.98	.48	.32	.23	.90	.27	.21
	(C1)	.33	.36	.44	.34	.39	.34		.50	.42	.44	.99	.49	.04	.35	.94	.40
	(X1)	.36	.33	.35	.38	.60	.43	.38		.39	.32	.51	.98	.11	.13	.32	.97
Incomplete Factor Scores	(S2)	.54	.46	.63	.66	.97	.47	.39	.57		.47	.44	.40	.89	.19	.25	.29
	(R2)	.45	.68	.36	.26	.47	.99	.33	.41	.45		.44	.30	.20	.92	.24	.19
	(C2)	.34	.27	.45	.35	.40	.34	.99	.28	.40	.34		.50	.06	.36	.94	.41
	(X2)	.36	.31	.34	.37	.59	.42	.36	.98	.57	.41	.36		.14	.10	.30	.98
Complete Factor Scores	(S3)	.48	.36	.57	.64	.91	.31	.12	.39	.93	.29	.13	.40		-.07	-.09	.05
	(R3)	.32	.63	.24	.08	.23	.93	.24	.21	.22	.95	.25	.20	.05		.17	.03
	(C3)	.15	.15	.28	.19	.14	.05	.93	.17	.13	.04	.93	.15	-.11	-.01		.21
	(X3)	.26	.20	.22	.25	.45	.26	.22	.95	.40	.23	.23	.96	.29	.08	.05	

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.

Table 1-6B: Intercorrelation Matrices for General, Composite and Factor Score Ratings; Sections E3 and E4

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
General Items	(GS)		.64	.60	.63	.76	.56	.40	.35	.76	.54	.41	.33	.68	.33	.20	.16
	(GR)	.55		.43	.51	.54	.64	.33	.21	.53	.64	.34	.22	.47	.51	.14	.06
	(GC)	.60	.49		.59	.59	.40	.50	.40	.56	.40	.51	.39	.47	.22	.35	.28
	(GX)	.52	.40	.66		.75	.50	.45	.30	.73	.48	.47	.29	.65	.27	.30	.14
Incomplete Factor Scores	(S1)	.70	.48	.74	.64		.57	.46	.36	.98	.54	.48	.34	.92	.23	.26	.14
	(R1)	.57	.59	.44	.46	.59		.36	.21	.54	.99	.37	.22	.46	.90	.10	.04
	(C1)	.51	.42	.56	.49	.56	.48		.56	.40	.36	.99	.57	.17	.18	.94	.44
	(X1)	.50	.44	.47	.44	.52	.47	.52		.29	.19	.57	.98	.10	.01	.44	.95
Incomplete Factor Scores	(S2)	.72	.50	.74	.63	.98	.59	.56	.48		.52	.42	.28	.96	.20	.20	.05
	(R2)	.56	.56	.42	.43	.57	.99	.47	.46	.56		.36	.20	.43	.92	.09	.01
	(C2)	.52	.42	.59	.50	.57	.49	.99	.53	.57	.47		.58	.19	.18	.93	.45
	(X2)	.49	.44	.45	.42	.50	.46	.50	.98	.46	.45	.51		.09	.02	.45	.95
Complete Factor Scores	(S3)	.64	.39	.63	.53	.90	.45	.29	.29	.94	.42	.31	.30		.11	-.02	-.12
	(R3)	.36	.43	.24	.26	.33	.91	.32	.25	.31	.93	.32	.23	.15		-.03	-.09
	(C3)	.34	.27	.44	.38	.37	.25	.94	.32	.36	.23	.94	.29	.11	.13		.36
	(X3)	.39	.37	.37	.35	.41	.34	.41	.97	.35	.33	.41	.97	.18	.37	.22	

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.

Table 1-7A (Part I): Squared Multiple Correlations and Beta Weights of General Ratings (GI) on Background Items, Pre- and Post-Course Motivation-Related Variables: Economics classes E1-E4

			(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
			E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
		R^2	.25*	.43@	.45@	.44@	.16	.32	.32*	.36@	.38@	.25	.34*	.60@	.31*	.43@	.32*	.38@
		$\ddot{R}^2$[2]	.11	.26	.31*	.32*	.01	.12	.15	.33*	.27*	.03	.18	.52@	.19	.27	.16	.35*
(B) Back-Ground	GPA	(2)[3]	-.03	-.04	-.09	-.15	.02	.06	-.01	-.02	.05	.11	.07	-.02	.08	-.04	-.05	-.02
	College Year	(5)	.06	-.03	.07	.15	.01	.04	-.08	.14	.03	.06	.05	-.18	.01	.02	- 03	-.03
	Major	(6)	.01	.21	.31@	.06	.03	.09	.39@	-.03	.20	.05	.18	-.29@	.18	.12	.11	-.14
	Optionality	(7)	-.08	.06	-.06	.08	-.16	-.09	-.10	.29*	-.37@	-.17	-.20	-.05	-.29@	-.21	-.07	-.10
	Class Size	(15)	.20	.15	.00	.14	.15	.24	.15	.15	.10	-.01	.10	.04	.07	.11	.04	.03
(P) Pre-Ratings	Amount Learned	(9)	.06	.38@	-.11	-.04	.10	.25	-.09	-.09	.10	.17	.00	-.12	.04	.42@	-.08	-.08
	Relevance	(10)	-.05	-.15	-.29	.16	-.05	-.33*	-.08	.13	.05	-.28	.03	.28@	-.14	-.36*	-.07	.02
	Workload	(12)	.04	.04	-.05	.05	.08	.13	-.07	.17	-.08	.13	-.08	.20*	-.02	.12	-.04	.13
	Interest	(13)	-.13	-.20	.31*	.10	-.10	-.08	.05	.16	-.15	-.09	.05	-.24*	-.03	.16	.07	-.06
	Effort	(14)	-.05	-.22	-.20	.00	.04	-.10	-.02	.13	.06	-.02	.10	-.10	.12	-.27*	.00	.02
(P) Post-Ratings	Amount Learned	(3)	.30@	.12	.34*	.32@	.06	-.03	.24	.35@	.41@	.16	.32@	.46@	.33@	.00	.24	.27*
	Relevance	(4)	.00	-.12	.16	-.23	.12	.06	.02	-.19	.03	.16	-.16	-.18	.12	.23	.14	-.06
	Workload	(5)	.14	-.13	-.04	-.18	.00	-.07	.01	-.18	.08	-.23	-.14	.04	.07	-.28*	.00	.06
	Interest	(6)	.27@	.39*	.13	.45@	.22	.20	.29	.22	.26@	.07	.36@	.44@	.22	.09	.33@	.47@
	Effort	(7)	-.23	.32*	-.08	-.09	-.12	.22	-.22	-.15	-.25@	.13	-.22	.07	-.16	.19	-.14	-.07

@ = $P \leq .01$
* = $P \leq .05$
1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---R^2 is the conventional notation for shrunken R2 (see Cohen and Cohen, 1975, p. 106-107).
3---Numbers in ( ) refer to the number in the original questionnaire.
Sample sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-7A: Squared Multiple and Semi-Partial Correlations of General Ratings (GI) on Background, Pre- and Post-Course Motivation-Related Variables; Economics classes E1-E4 (Part II)

	(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.B}$[2]	.05	.07	.11	.04	.05	.13	.13	.06	.09	.08	.06	.09	.07	.05	.26	.04
$\tilde{R}^2_{y.B}$[3]	.00	.00	.00	.00	.00	.77	.07	.00	.01	.01	.00	.03	.00	.00	.00	.00
$R^2_{y.PB}$	.08	.22	.26	.17	.08	.24	.16	.19	.13	.14	.14	.20	.13	.30*	.10	.12
$\tilde{R}^2_{y.PB}$	.00	.10	.15	.15	.00	.13	.04	.16	.03	.01	.01	.18	.03	.20	.00	.10
$R^2_{y.\underline{P}PB}$	.25*	.43@	.45@	.44@	.16	.32	.32*	.36@	.38@	.25	.34*	.60@	.31*	.43@	.32*	.38@
$\tilde{R}^2_{y.\underline{P}PB}$	.11	.26	.31*	.32*	.01	.12	.15	.33*	.27*	.03	.18	.52@	.19	.27	.16	.35*
$R^2_{y.PB}-R^2_{y.B}$	.03	.15	.14	.12	.03	.11	.04	.13	.04	.06	.08	.11	.06	.25@	.04	.09
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.00	.10	.15	.15	.00	.12	-.03	.16	.02	.00	.01	.15	.03	.20*	.00	.10
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.17@	.21*	.19*	.27@	.08	.07	.16	.17*	.25@	.11	.21@	.40@	.20@	.13	.20@	.26@
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.11	.15	.16*	.17*	.01	.00	.11	.17*	.25@	.00	.17*	.34@	.16*	.07	.16*	.26*

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; B=background items; P=Pre-course motivation-related item; $\underline{P}$=Post-Course motivation-related item.
3---R^2 is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-7B (Part I): Squared Multiple Correlations and Beta Weights of Composite Ratings (IFS) on Background Items, Pre- and Post-Course Motivation-Related Variables

			(S1)[1] Skill				(R1) Rapport				(C2) Content				(X2) Organization/ Structure			
			E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation		$R^2_{y.\underline{P}PB}$[3]	.34@	.35*	.33*	.45@	.23	.34	.26	.34@	.53@	.59@	.64@	.60@	.34@	.23	.48@	.49@
Adjusted R^2		$\tilde{R}^2_{y.\underline{P}PB}$[5]	.22	.16	.17	.33@	.09	.14	.08	.20	.45@	.47*	.55@	.53@	.33	.02	.36@	.38@
(B) Background	GPA	(2)[4]	.06	.08	-.14	.09	.03	.02	-.10	.00	-.09	-.01	.17	-.12	-.26*	-.10	.02	-.15
	College Year	(5)	.09	.05	.12	.14	.06	.09	.01	.24*	.05	.03	.02	.06	.18	-.10	.02	.31*
	Major	(6)	.16	.04	.09	-.21	.06	.17	.21	-.06	.05	.03	.06	.03	.08	.10	-.16	.01
	Optionality	(7)	-.24	-.17	.01	.04	.02	.20	-.11	.13	-.16	-.07	-.16	-.11	-.19	-.06	.23	.09
	Class Size	(15)	.09	.14	-.01	.15	.06	.08	-.10	.08	.02	.10	-.04	.08	.12	-.02	.02	.06
	Amount Learned	(9)	.17	.28	.09	.10	.02	.09	.00	.05	.08	.09	.31*	-.05	.07	.06	.18	.06
($\underline{P}$) Pre-Ratings	Relevance	(10)	-.15*	-.13	-.21	.23*	-.18	-.17	-.24	.11	.00	-.29	.16	.05	-.04	-.21	-.22	.13
	Workload	(12)	.09	.07	-.06	.06	.06	.03	.20	.04	.04	.10	-.01	.15	-.02	-.01	-.04	.01
	Interest	(13)	-.15	-.03	.20	.07	.18*	-.27	.12	.10	.22	.00	-.08	-.01	.02	.06	.40*	.29*
	Effort	(14)	.02	-.21	-.20	.12	.14	-.17	.21	.12	-.03	-.11	-.01	-.04	-.09	.11	-.07	-.04
($\underline{P}$) Post-Ratings	Amount Learned	(3)	.24*	.21	.23	.36@	.20	-.10	.19	.17	.15	-.02	.32@	.34@	.27@	.21	.49@	.50@
	Relevance	(4)	.02	.16	.19	-.23	-.04	.11	.22	-.21	.17	.43@	-.01	.26*	.10	.16	.39*	-.15
	Workload	(5)	-.11	-.40*	.00	-.10	-.04	.06	-.11	-.01	-.12	-.20	-.13	.03	-.24*	-.06	-.07	-.11
	Interest	(6)	.30*	-.02	.20	.42@	.17	.37	.05	.46@	.35@	.38*	.43@	.35@	.16	.04	-.36*	.17
	Effort	(7)	-.17	.34*	-.18	-.21	-.12	.35	.01	-.24	.16	.26*	-.04	-.05	.01	.28	.07	-.35*

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; P=pre-course motivation-related item; $\underline{P}$=post-course motivation-related item; B=background variable.
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
5---R^2 is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-7B (Part II): Squared Multiple and Semi-Partial Correlations of Composite Ratings (IC) on Background, Pre- and Post-Course Motivation-Related Variables: Economics E1-E4

	(S1)[1] Skill				(R1) Rapport				(C1) Content				(X1) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.B}$[3]	.06	.06	.07	.04	.02	.08	.08	.03	.04	.07	.09	.01	.10	.02	.01	.03
$\tilde{R}^2_{y.B}$[4]	.01	.00	.01	.00	.00	.01	.02	.00	.00	.00	.03	.00	.05	.00	.00	.00
$R^2_{y.PB}$	.15	.14	.17	.18	.16	.12	.16	.16	.26@	.16	.34@	.25@	.16	.09	.18	.20*
$\tilde{R}^2_{y.PB}$	.05	.00	.05	.07	.06	.00	.03	.04	.17	.01	.24*	.15	.06	.00	.06	.10
$R^2_{y.\underline{P}PB}$	.34@	.35*	.33*	.45@	.23	.34	.26	.34@	.52@	.57@@	.67@	.59@	.33@	.25	.47@	.46@
$\tilde{R}^2_{y.\underline{P}PB}$	.22	.16	.17	.33@	.09	.14	.08	.20	.43@	.45@	.59@	.51@	.20	.03	.35*	.34@
$R^2_{y.PB}-R^2_{y.B}$	.09	.09	.10	.14*	.14*	.04	.08	.12	.22@	.09	.25@	.24@	.06	.07	.18*	.18@
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.04	.00	.04	.07	.06	.00	.01	.04	.17@	.01	.22@	.15*	.01	.00	.06	.10
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.19@	.20*	.16	.27@	.07	.22@	.10	.19*	.27@	.42@	.32@	.34@	.17*	.16	.30@	.25@
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.17@	.16	.12	.26@	.03	.14	.15	.16*	.26@	.44@	.34@	.35@	.14*	.03	.29@	.24@

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; P=post-course motivation-related item; P=pre-course motivation-related item; B= background variable.
4---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = P ≤ .01
* = P ≤ .05
Sample sizes: N_{E1}=96, N_{E2}=68, N_{E3}=79, N_{E4}=87.

Table 1-7C (Part I): Squared Multiple Correlations and Beta Weights of Composite Ratings (IC) on Background Items, Pre- and Post-Course Motivation-Related Variables

			(S2)[1] Skill				(R2) Rapport				(C1) Content				(X1) Organization/ Structure			
			E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation		$R^2_{y.PPB}$[3]	.32@	.36*	.30	.45@	.23	.36*	.23	.29*	.52@	.57@	.67@	.59@	.33@	.25	.47@	.46@
Adjusted R^2		$\tilde{R}^2_{y.PPB}$[5]	.20	.18	.13	.33@	.09	.18	.04	.14	.43@	.45@	.59@	.51@	.20	.03	.35@	.34@
(B) Background	GPA	(2)[4]	.07	.11	-.14	.11	.04	.13	-.06	-.01	-.10	.00	.19*	-.13	-.25*	-.09	.00	-.14
	College Year	(5)	.07	.06	.12	.12	.05	.07	-.01	.19	.05	.03	.02	.07	.15	-.08	.00	.25
	Major	(6)	.11	.28	.14	-.23	.03	.17	.24	.01	.05	.04	.09	.03	.07	.07	-.21	.05
	Optionality	(7)	-.21	-.16	-.05	.08	.03	.24	-.15	.10	-.15	-.08	.20	-.11	-.17	-.06	.30	.04
	Class Size	(15)	.06	.16	.02	.19	.08	.06	-.08	.08	.02	.11	-.03	.07	.13	.01	.02	.08
(P) Pre-Ratings	Amount Learned	(9)	.13	.29*	.08	-.13	.02	.04	.02	.05	.06	.08	.32@	-.05	.07	.08	.18	.16
	Relevance	(10)	-.19	-.16	-.17	.24	.08	-.11	-.19	.08	.01	-.30	.21	.04	-.05	-.30	-.26	.12
	Workload	(12)	.02	.01	-.07	.08	.06	.02	.16	.02	.04	.10	.00	.15	.02	.10	-.06	.01
	Interest	(13)	-.13	-.08	.15	.07	.17	-.24	.09	.05	.22	-.01	-.12	-.18	.05	.10	.37*	.24
	Effort	(14)	.08	-.17	-.16	.11	.14	-.19	-.17	.11	-.03	-.10	.02	-.04	-.10	.01	-.12	-.10
(P) Post-Ratings	Amount Learned	(3)	.26*	.19	.21	.32@	.22	-.10	.20	.17	.14	-.03	.31	.33@	.24*	.20	.50@	.51
	Relevance	(4)	.06	.18	.15	-.26*	-.07	.07	.14	-.18	.20	.48@	-.03@	.30@	.08	.18	.42@	-.13
	Workload	(5)	-.06	-.42@	-.01	-.12	-.02	.08	-.10	-.01	-.11	-.20	-.15	.03	-.21	-.11	-.05	-.10
	Interest	(6)	.29*	-.01	.20	.44@	.15	.39*	.06	.43@	.34@	.35*	.45@	.35@	.19	.00	-.30*	.18
	Effort	(7)	-.20	.34*	-.07	-.14	-.13	.36*	-.01	-.19	.16	.25*	-.06	-.05	.12	.31	.04	-.24

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y criterion variable; P=pre-course motivation-related item; B=background variable.
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
5---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: N_{E1}=96, N_{E2}=68, N_{E3}=79, N_{E4}=87.

Table 1-7C (Part II): Squared Multiple and Semi-Partial Correlation of Composite Ratings (IFS) on Pre- and Post-Course Motivation-Related Variables: Economics E1-E4

	(S2)[1] Skill				(R2) Rapport				(C2) Content				(X2) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.B}$[3]	.04	.06	.09	.05	.02	.09	.09	.04	.04	.08	.08	.01	.10	.02	.01	.04
$\tilde{R}^2_{y.B}$[4]	.00	.00	.02	.00	.00	.01	.02	.00	.00	.000	.02	.00	.05	.00	.00	.00
$R^2_{y.PB}$	.13	.14	.16	.19	.16	.12	.14	.13	.27@	.17	.32@	.26@	.14	.07	.21	.23*
$\tilde{R}^2_{y.PB}$	.03	.00	.00	.08	.06	.00	.02	.02	.18	.02	.22	.16	.04	.00	.09	.13
$R^2_{y.\underline{P}PB}$	.32@	.36*	.30	.45@	.23	.36@	.23	.29*	.53@	.59@	.64@	.60@	.34@	.23	.48@	.49*
$\tilde{R}^2_{y.\underline{P}PB}$	.20	.18	.13	.33*	.09	.18	.04	.14	.45@	.47*	.55@	.52@	.22	.01	.36*	.38*
$R^2_{y.PB}-R^2_{y.B}$	.09	.08	.07	.14	.14	.03	.05	.09	.23	.09	.24	.25	.04	.05	.20	.19
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.03	.00	.00	.08	.06	.00	.00	.02	.18@	.02	.20@	.16*	.00	.00	.09	.09
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.19@	.22@	.14	.26@	.07	.24@	.21	.16@	.26@	.42@	.32@	.34@	.20@	.16	.27@	.26@
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.17@	.18	.13	.15@	.03	.18	.02	.12	.27@	.45@	.33@	.36@	.18@	.01	.27@	.25@

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; P=pre-course motivation-related item; B=background variable. P=post-course motivation-related item;
4---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample Sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-7D (Part I): Squared Multiple Correlations and Bet a Weights of Composite Ratings (CFS) on Background Items, Pre- and Post-Cours e Motivation-Related Variables

		(S3)[1] Skill				(R3) Rapport				(C3) Content				(X3) Organization/ Structure			
		E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation	$R^2_{y.\underline{P}PB}$[3]	.22	.36*	.21	.37	.18	.31	.15	.21	.49@	.49@	.59@	.56@	.31@	.16	.40@	.44@
Adjusted R^2	$\tilde{R}^2_{y.\underline{P}PB}$[5]	.05	.18	.02	.24	.03	.11	.00	.04	.39	.34	.49	.47	.18	.00	.26	.32
(B) Back-Ground: GPA	(2)[4]	.14	.14	-.16	.14	.03	.08	-.04	-.03	-.05	-.08	.23	-.12	-.29	-.05	-.02	-.18
(B) Back-Ground: College Year	(5)	.05	.16	.17	.09	-.03	.06	-.11	.15	.02	.01	.03	.00	.17	-.05	-.02	.27*
(B) Back-Ground: Major	(6)	.10	-.05	.17	-.31	-.01	.18	.20	.12	.07	-.01	.03	.03	.07	-.02	-.32	.06
(B) Back-Ground: Optionality	(7)	-.15	-.18	-.04	.05	.07	.30	-.15	.04	-.15	.11	-.13	-.15	-.16	-.04	.36*	.05
(B) Back-Ground: Class Size	(15)	-.04	.22	.04	.16	.12	.05	-.13	.02	-.03	.03	-.04	.05	.15	.06	.03	.00
(P) Pre-Ratings: Amount Learned	(9)	.14	.36@	.04	-.15	.00	-.02	-.08	.16	.04	.02	.32@	-.05	.07	.08	.09	.07
(P) Pre-Ratings: Relevance	(10)	-.15	-.12	-.11	.29*	.04	-.12	-.26	-.08	.00	-.24	.29*	.01	-.03	-.23	-.30	.10
(P) Pre-Ratings: Workload	(12)	-.05	.11	.08	.01	.09	.00	.20	-.03	.04	.09	-.04	.20	.00	-.05	-.02	-.01
(P) Pre-Ratings: Interest	(13)	-.23	-.12	.10	.06	.21	-.23	.10	-.05	.24@	.09	-.18	-.05	.00	.18	.44@	.28*
(P) Pre-Ratings: Effort	(14)	.11	-.17	-.17	.12	.17	-.15	-.10	.16	-.08	-.07	.06	-.08	-.05	.05	-.10	-.08
(P) Post-Ratings: Amount Learned	(3)	.19	.21	.15	.22	.08	-.17	.09	.00	.08	-.05	.25	.26*	.22*	.26	.42	.52@
(P) Post-Ratings: Relevance	(4)	-.04	.08	.06	-.40@	-.14	.14	.15	-.09	.25@	.42@	-.13	.41@	.05	.13	.47@	-.12
(P) Post-Ratings: Workload	(5)	.09	-.45@	.03	-.15	.03	.15	-.09	.08	-.10	-.17	-.14	.06	-.24*	-.09	.00	-.08
(P) Post-Ratings: Interest	(6)	.21	-.13	.18	-.39@	.08	.35	-.03	.40@	.30@	.34@	.53@	.25*	-.15	-.18	-.46@	.06
(P) Post-Ratings: Effort	(7)	-.25*	.26	-.10	-.08	-.07	.32	.09	-.24	.20*	.15	-.08	.04	.06	.20	.13	-.34*

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; $\underline{P}$=post-course motivation-related item; P=pre-course motivation-related item; B= background item.
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
5---R^2 is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-7D (Part II): Squared Multiple and Semi-Partial Correlations of Composite Ratings (CFS) on Background, Pre- and Post-Course Motivation-Related Variables: Economics E1-E4

	(S3)[1] Skill				(R3) Rapport				(C3) Content				(X4) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.B}$[3]	.02	.08	.10	.08	.02	.11	.07	.06	.04	.04	.06	.01	.13*	.01	.01	.05
$\tilde{R}^2_{y.B}$[4]	.00	.00	.04	.02	.00	.04	.01	.00	.00	.00	.00	.00	.08	.00	.00	.00
$R^2_{y.PB}$	.11	.21	.15	.16	.16	.14	.12	.10	.23@	.17	.29@	.24*	.17	.06	.14	.21*
$\tilde{R}^2_{y.PB}$	.01	.17	.03	.05	.06	.00	.00	.00	.14	.02	.19	.14	.07	.00	.01	.11
$R^2_{y.\underline{P}PB}$	.22	.36*	.21	.37@	.18	.31	.15	.21	.49@	.49@	.59@	.56@	.31@	.16	.40@	.44@
$\tilde{R}^2_{y.\underline{P}PB}$	.05	.18	.02	.24	.03	.11	.00	.04	.39	.34	.49	.47	.18	.00	.26	.32
$R^2_{y.PB}-R^2_{y.B}$	.09	.13	.05	.08	.14	.03	.05	.04	.19	.13	.23	.23	.04	.05	.13	.16
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.01	.07	.00	.03	.06	.00	.00	.00	.14	.02	.19	.14	.00	.00	.01	.11
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.11	.15	.06	.21	.02	.17	.03	.11	.26@	.32@	.30@	.32@	.14	.10	.26@	.23@
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.14	.11	.00	.19@	.00	.00	.00	.04	.25@	.32@	.30@	.33@	.11	.00	.25@	.21@

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; B=background variable; P=pre-course motivation-related item; $\underline{P}$=post-course motivation-related item.
4---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample Sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-8: Summary of Adjusted Squared Multiple and Semi-Partial Correlations for Tables 1-7A through D; Economics classes E1-E4

		Skill				Rapport				Content				Organization/ Structure			
$\tilde{R}^2_{y.\underline{P}PB}$[2]		E1[1]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
	GI[4]	.11	.26	.31*	.32*	.01	.12	.15	.33*	.27*	.03	.18	.52@	.19	.27	.16	.35@
	IC	.22	.16	.17	.33@	.09	.14	.08	.20	.43@	.45@	.59@	.51@	.20	.03	.35@	.34@
	IFS	.20	.18	.13	.33@	.09	.18	.04	.14	.45@	.47@	.55@	.52@	.22	.01	.36@	.38@
	CFS	.05	.18	.02	.24	.03	.11	.00	.04	.39	.34	.49	.47	.18	.00	.26	.32
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$																	
	GI	.00	.10	.15*	.15*	.00	.12	.00	.16*	.02	.00	.01	.15*	.03	.20*	.00	.10
	IC	.04	.00	.04	.07	.06	.00	.01	.04	.17@	.01	.22@	.15*	.01	.00	.06	.10
	IFS	.03	.00	.00	.08	.06	.00	.00	.02	.18@	.02	.20@	.16*	.00	.00	.09	.09
	CFS	.01	.07	.00	.03	.06	.00	.00	.00	.14*	.02	.19@	.14*	.00	.00	.01	.11
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$																	
	GI	.11	.15	.16*	.17@	.01	.00	.11	.17@	.25@	.02	.17*	.34@	.16*	.07	.16*	.26@
	IC	.17@	.16	.12	.26@	.03	.14	.05	.16*	.26@	.44@	.34@	.35@	.14*	.03	.29@	.24@
	IFS	.17@	.18	.13	.15@	.03	.18	.02	.12	.27@	.45@	.33@	.36@	.18@	.01	.27@	.25@
	CFS	.04	.11	.00	.19@	.00	.00	.00	.04	.25@	.32@	..30@	.33@	.11	.00	.25@	.21@

1---Letter and number symbols refer to discipline (economics) and section.
2---Subscript notation: y=criterion variable; $\underline{P}$=post-course motivation-related item; P=pre-course motivation-related item; B= background variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
4---GI=general rating items; IC=incomplete (equally weighted) composite scores; IFS=incomplete (optimally weighted) factor scores; CFS=complete (optimally weighted) factor scores.

Table 1-9: Squared Multiple Correlations, Squared Semi-Partial Correlations and Beta Weights from Regression of Pre-Course Motivtion-Related Variables on Background and Related Items; Economics classes E1-E4

	(11) Amount Learned				(12) Relevance				(14) Workload				(15) Interest				(16) Effort			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
GPA (2)[1]	.10	.01	-.08	-.07	.15	.12	.11	.00	.12	.07	-.03	-.04	.12	.12	-.02	.02	.14	.06	-.13	-.10
College Year (3)	.10	.04	.00	-.01	-.02	-.02	.05	-.05	.06	-.01	.09	-.04	.07	.13	.04	-.05	.15	.18	.15	.10
Major (5)	.03	.00	.28*	.00	.20	.19	.29*	.10	.05	.20	.18	.24	-.03	.01	.13	-.06	.14	.24	.21	.19
Optionality (6)	.14	.16	-.18	.10	.07	.20	.05	.11	-.15	-.06	-.15	-.13	-.12	.03	-.25*	-.08	-.07	.08	.02	.04
Class Size (15)	-.09	-.04	-.03	-.05	.00	-.06	.04	.00	-.08	.02	.00	-.05	.11	.02	.08	.01	-.10	.02	-.16	-.02
Like Soc/Sci (18)	.13	.23	.27*	.22	.27@	.26*	.30*	.36@	.08	.27*	.25*	.18	.30@	.36@	.35@	.40@	.28@	.14	.24	.26*
Prior Grade (22)	-.32	-.06	-.04	-.09	-.15	.07	-.21	-.06	.00	-.12	-.08	.02	-.19	.03	.00	-.04	-.14	-.17	-.04	-.21
Prior Rating (23)	.16	-.13	-.13	.00	.20	-.16	.11	.00	.04	.20	.00	-.08	.18	.11	.11	.10	.21	.10	-.07	.13
$R^2_{y.B}$[2]	.02	.01	.06	.02	.08	.10	.12	.04	.04	.07	.05	.05	.07	.05	.09	.03	.09	.11	.10	.10
$\tilde{R}^2_{y.B}$[3]	.00	.00	.00	.00	.03	.03	.04	.00	.00	.00	.00	.00	.02	.00	.01	.00	.02	.05	.03	.04
$R^2_{y.CB}$	.08	.08	.14	.07	.17*	.17	.21*	.16	.05	.17	.11	.08	.17*	.21	.23@	.20*	.19*	.14*	.16	.18*
$\tilde{R}^2_{y.CB}$	.00	.00	.04	.00	.09	.06	.12	.06	.00	.05	.01	.00	.10	.10	.14	.11	.12	.02	.05	.09
$R^2_{y.CB}-R^2_{y.B}$	.06	.06	.08	.05	.09	.07	.09	.12	.01	.10	.06	.04	.11	.16*	.14	.17	.10	.03	.06	.09
$\tilde{R}^2_{y.CB}-\tilde{R}^2_{y.B}$	.00	.00	.04	.00	.07	.05	.07	.06	.00	.05	.01	.00	.08	.10	.13	.11	.08	.00	.02	.00

@ = $p \leq .01$
* = $p \leq .05$
1---Numbers in ( ) refer to the variable numbers in the original questionnaire.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
3---Subscript notation: y=criterion variable; B= background characteristics;
C=equally weighted composite variable.
Sample sizes: $N_{E1}=96$, $N_{E2}=68$, $N_{E3}=79$, $N_{E4}=87$.

Table 1-10A (Part I): Squared Multiple Correlations and Be ta Weights of General Ratings on Composite Ratings (IC) and Post-Course Items; Economics classes E1-E4

		(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation	$R^2_{y.IPC}$[2]	.56@	.46@	.74@	.67@	.43@	.64@	.58@	.48@	.49@	.46@	.50@	.63@	.49@	.47@	.63@	.49@
Adjusted R^2	$\tilde{R}^2_{y.IPC}$[3]	.53@	.39@	.71@	.64@	.39@	.59@	.54@	.44@	.47@	.41@	.45@	.60@	.46@	.40@	.59@	.45@
Composite Scores: Skill	(S1)	.32@	.24*	.43@	.24@	.21@	.11	.04	-.07	.35@	.51@	.27@	.61@	.39@	.59@	.60@	.40@
Composite Scores: Rapport	(R1)	.04	.20*	.02	.12	.20@	.55@	.37@	.40@	-.12	.05	-.02	-.08	.03	-.12	.07	.01
Composite Scores: Content	(C1)	.02	.05	.11	.03	.19*	.17	-.08	.02	-.03	.29@	.18	.02	.08	.00	.06	-.02
Composite Scores: Organization/ Structure	(X1)	.02	-.04	.09	.06	.03	-.11	-.02	.07	.05	-.10	.14	-.02	.07	-.01	.02	.07
Post-Course Ratings: Amount Learned	(3)[4]	.13*	.18	.16*	.07	-.10	.01	.10	.14	.32@	.03	.05	.23@	.18@	-.12	.02	.05
Post-Course Ratings: Relevance	(4)	-.03	-.16	-.05	-.07	.02	-.23@	-.01	.02	.03	-.13	-.13	.01	.01	.02	.07	.00
Post-Course Ratings: Workload	(5)	.10	-.03	-.02	-.09	-.03	.15	.17*	-.13	.07	.10	-.05	-.01	.10	.07	.02	.04
Post-Course Ratings: Interest	(6)	-.03	.15	.04	.09	-.08	.07	.10	-.05	.06	.04	.18	.27@	.01	.15	.19*	.27@
Post-Course Ratings: Effort	(7)	-.04	.08	-.02	.08	.03	-.11	-.14	.07	-.06	-.09	-.14	-.08	.00	-.05	-.06	-.04
Post-Course Ratings: Enthusiasm	(8)	.34@	.08	.16*	.18*	.28@	.16	.30@	.23@	.24@	.02	.11	.01	.22	.11	.18*	.03
Post-Course Ratings: Entertainment	(9)	.22@	.18	.33@	.30@	.11	.18*	.15	.17*	.12	.02	.14	.01	-.02	.14	-.04	.11

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; C=equally weighted composite variable; P=post-course motivation-related item; I= incidental variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$

Table 1-10A (Part II): Squared Multiple and Semi-Partial Correlations from Regression of General Items on Composite Ratings (IC), Post-Course Motivation-Related Items, and Teacher Enthusiasm and Entertainment

	(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.C}$[2]	.36@	.33@	.60@	.54@	.34@	.53@	.46@	.40@	.33@	.43@	.42@	.59@	.41@	.41@	.58@	.44@
$\tilde{R}^2_{y.C}$[3]	.35@	.30@	.59@	.53@	.33@	.51@	.44@	.38@	.32@	.41@	.40@	.58@	.40@	.39@	.57@	.42@
$R^2_{y.\underline{P}C}$	.41@	.42@	.64@	.57@	.35@	.59@	.49@	.41@	.43@	.45@	.47@	.63@	.48@	.45@	.61@	.48@
$\tilde{R}^2_{y.\underline{P}C}$	.38@	.36@	.61@	.54@	.32@	.55@	.45@	.37@	.40@	.39@	.43@	.60@	.45@	.39@	.58@	.45@
$R^2_{y.I\underline{P}C}$	.56@	.46@	.74@	.67@	.43@	.64@	.58@	.48@	.49@	.46@	.50@	.63@	.49@	.47@	.63@	.49@
$\tilde{R}^2_{y.I\underline{P}C}$	.53@	.39@	.71@	.64@	.39@	.59@	.54@	.44@	.47@	.41@	.45@	.60@	.46@	.40@	.59@	..45@
$R^2_{y.\underline{P}C}-R^2_{y.C}$	.04	.09	.04	.02	.01	.06	.03	.01	.10	.02	.05	.04	.07	.04	.03	.04
$\tilde{R}^2_{y.\underline{P}C}-\tilde{R}^2_{y.C}$	.03	.06	.02	.01	.00	.04	.01	.00	.08	.00	.03	.02	.05	.00	.01	.03
$R^2_{y.I\underline{P}C}-R^2_{y.\underline{P}C}$	.16@	.04*	.10@	.10@	.08@	.05**	.09@	.07@	.06@	.01	.03	.00	.01	.02	.02	.01
$\tilde{R}^2_{y.I\underline{P}C}-\tilde{R}^2_{y.\underline{P}C}$	.15@	.03	.10@	.10@	.07@	.04*	.09@	.07@	.07@	.00	.02	.00	.01	.01	.01	.00

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; C=equally weighted composite variable; P=post-course motivation-related variable; I=incidental variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$

Table 1-10B (Part I): Squared Multiple Correlations and Be ta Weights of General Ratings on Composite Ratings (IFS) and Post-Course Items; Economics class E1-E4

		(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation	$R^2_{y.IPF}$[2]	.56@	.46@	.75@	.69@	.44@	.63@	.59@	.47@	.47@	.47@	.50@	.64@	.49@	.49@	.61@	.49@
Adjusted R^2	$\tilde{R}^2_{y.IPF}$[3]	.53@	.39@	.73@	.66@	.40@	.58@	.55@	.43@	.44@	.40@	.45@	.60@	.45@	.43@	.57@	.45@
Factor Scores: Skill	(S2)	.30@	.24*	.44@	.30@	.21@	.07	.07	.02	.27@	.53@	.26@	.59@	.38@	.63@	.54@	.40@
Factor Scores: Rapport	(R2)	.04	.19	.04	.12	.21@	.54@	.37@	.34@	-.07	.06	.02	-.07	.04	-.13	.08	-.01
Factor Scores: Content	(C2)	.01	.05	-.09	.02	.19*	.20*	-.10	-.02	-.03	.30@	.19	.02	.08	.02	.00	-.02
Factor Scores: Organization/ Structure	(X2)	.05	-.01	.09	.08	.03	-.10	-.01	.09	.06	-.09	.16	.02	.09	-.01	.05	.06
Post-Course Ratings: Amount Learned	(3)[4]	.11	.18	.17*	.07	-.11	.01	.09	.13	.30	.03	.05	.26@	.17*	-.12	.02	.07
Post-Course Ratings: Relevance	(4)	-.04	-.17	-.07	.07	.03	-.23*	.00	.03	.02	-.13	-.14	.01	.00	.01	.05	.00
Post-Course Ratings: Workload	(5)	.08	-.03	-.02	-.07	-.04	.14	.16	-.12	.05	.12	-.05	.00	.08	.09	.02	.05
Post-Course Ratings: Interest	(6)	-.03	.13	.04	.08	-.08	.04	.10	-.03	.07	.01	.20	.24@	.01	.12	.19*	.26@
Post-Course Ratings: Effort	(7)	-.03	.06	-.04	.07	.03	-.13	-.14	.07	-.06	-.11	-.15	-.11	.01	-.05	-.08	-.06
Post-Course Ratings: Enthusiasm	(8)	.34@	.08	.16*	.16*	.28@	.18*	.30@	.23*	.25@	.02	.10	.02	.23@	.11	-.08	.04
Post-Course Ratings: Entertainment	(9)	.22@	.17	.31@	.29@	.10	.20*	.15	.16*	.12	.08	.13	.02	-.03	.12	.17	.12

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; F=incomplete factor score composite variable; P=post -course motivation-related item; I= incidental variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-10B (Part II): Squared Multiple and Semi-Partial Correlations and Beta Weights of General Ratings on Composite Ratings (IFS) and Post-Course Items; Economics classes E1-E4

	(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.F}$[3]	.37@	.35@	.62@	.57@	.35@	.51@	.46@	.39@	.30@	.45@	.42@	.58@	.42@	.44@	.57@	.44@
$\tilde{R}^2_{y.F}$[2]	.36@	.32@	.61@	.56@	.34@	.49@	.44@	.37@	.28@	.43@	.40@	.57@	.41@	.42@	.56@	.42@
$R^2_{y.\underline{P}F}$	.40@	.43@	.66@	.59@	.36@	.57@	.50@	.40@	.40@	.47@	.48@	.63@	.46@	.47@	.60@	.48@
$\tilde{R}^2_{y.\underline{P}F}$	.37@	.41@	.65@	.56@	.33@	.53@	.46@	.36@	.37@	.42@	.44@	.61@	.43@	.42@	.57@	.45@
$R^2_{y.I\underline{P}F}$	.56@	.46@	.75@	.69@	.44@	.63@	.59@	.47@	.47@	.47@	.50@	.64@	.49@	.49@	.61@	.49@
$\tilde{R}^2_{y.I\underline{P}F}$	.53@	.39@	.73@	.66@	.40@	.58@	.55@	.43@	.44@	.40@	.45@	.60@	.45@	.43@	.57@	.45@
$R^2_{y.\underline{P}F}-R^2_{y.F}$	.03	.08	.04	.02	.01	.06	.04	.01	.10@	.02	.06	.05	.04	.03	.03	.04
$\tilde{R}^2_{y.\underline{P}F}-\tilde{R}^2_{y.F}$	.01	.09	.04	.00	.00	.04	.02	.00	.09@	.00	.04	.04	.02	.00	.01	.03
$R^2_{y.I\underline{P}F}-R^2_{y.\underline{P}F}$	.16@	.03	.09@	.10@	.08@	.06*	.09	.07@	.07@	.00	.02	.01	.03	.02	.01	.01
$\tilde{R}^2_{y.I\underline{P}F}-\tilde{R}^2_{y.\underline{P}F}$	.16@	.00	.08@	.10@	.07@	.05*	.09@	.07@	.07@	.00	.01	.00	.02	.01	.00	.00

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
3---Subscript notation: y=criterion variable; I=incidental variables; P=pre-course motivation-related; $\underline{P}$=post-course motivation-related; F=incomplete factor score composite variable.
* = $P \leq .05$
@ = $P \leq .01$
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-10C (Part I): Squared Multiple Correlations and Be ta Weights of General Ratings on Composite Ratings (CFS) and Post-Course Items; Economics classes E1-E4

		(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Squared Multiple Correlation	$R^2_{y.IPF}$[2]	.56@	.48@	.47@	.75@	.44@	.64@	.59@	.46@	.48@	.51@	.50@	.63@	.50@	.52@	.62@	.49@
Adjusted R^2	$\tilde{R}^2_{y.IPF}$[3]	.53@	.40@	.70@	.66@	.40@	.59@	.55@	.42@	.45@	.45@	.45@	.60@	.47@	.46@	.58@	.45@
Factor Scores: Skill	(S3)	.29@	.29@	.42@	.33@	.28@	.22@	.17*	.10	.25@	.61@	.35@	.48@	.41@	.63@	.58@	.33@
Factor Scores: Rapport	(R3)	.10	.22*	.09	.14*	.28@	.56@	.31@	.30@	-.06	.24@	.11	.04	.11*	.00	.17*	.05
Factor Scores: Content	(C3)	.10	.12	.00	.08	.24@	.22	-.02	.01	.09	.42@	.16	.15	.19@	.20	.08	.11
Factor Scores: Organization/ Structure	(X3)	.11	.05	.14	.12	.13	.03	-.01	.13	.10	.02	.27@	.06	.18@	.06	.17	.13
Post-Course Ratings: Amount Learned	(3)[4]	.12	.17	.16	.07	-.10	.00	.12	.13	.29	.03	.02	.25@	.17@	-.13	-.01	.05
Post-Course Ratings: Relevance	(4)	-.03	-.16	-.06	-.04	.05	-.21@	.12	.05	.01	-.13	-.13	.02	.01	.00	.05	.00
Post-Course Ratings: Workload	(5)	.07	-.01	-.03	-.08	-.04	.17	.15*	-.14	.05	.14	-.05	.00	.08	.11	.02	.05
Post-Course Ratings: Interest	(6)	-.04	.13	.06	.08	-.09	.05	.09	-.03	.04	.00	.25*	.22@	-.01	.09	.23*	.25
Post-Course Ratings: Effort	(7)	-.04	.05	-.04	.08	.03	-.16	-.16	.09	-.06	-.16	-.14	-.12	.00	-.09	-.08	-.06
Post-Course Ratings: Enthusiasm	(8)	.34@	.05	.15	.15*	.28@	.14	.31@	.22*	.25@	-.03	.07	.01	.21@	.05	-.12	.04
Post-Course Ratings: Entertainment	(9)	.22@	.18*	.31@	.28@	.09	.23@	.12	.16	.11	.07	.14	.01	-.03	.10	.17*	.12

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; F=complete factor score composite variable; P=post-course motivation-related item; I= incidental variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
4---Numbers in ( ) refer to the variable numbers in the original questionnaire.
Sample Sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-10C (Part II): Squared Multiple and Semi-Partial Correlations for Regression of General Items on Composite Ratings (CFS), Post-Course Motivation-Related Items, and Teacher Enthusiasm and Entertainment

	(GS)[1] Skill				(GR) Rapport				(GC) Content				(GX) Organization/ Structure			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.\underline{F}}$[2]	.37@	.36@	.62@	.58@	.34@	.55@	.47@	.38@	.33@	.48@	.43@	.58@	.43@	.49@	.58@	.44@
$\tilde{R}^2_{y.\underline{F}}$[3]	.36@	.33@	.61@	.57@	.33@	.53@	.45@	.36@	.32@	.46@	.41@	.57@	.42@	.47@	.57@	.42@
$R^2_{y.\underline{PF}}$	.40@	.44@	.65@	.60@	.35@	.59@	.50@	.40@	.42	.50@	.48@	.62@	.47@	.51@	.60@	.47@
$\tilde{R}^2_{y.\underline{PF}}$	.37@	.38@	.62@	.57@	.32@	.55@	.46@	.36@	.39@	.45@	.44@	.59@	.44@	.46@	.57@	.44@
$R^2_{y.I\underline{PF}}$	.56@	.47@	.73@	.69@	.44@	.64@	.59@	.46@	.48@	.51@	.50@	.63@	.50@	.52@	.62@	.49@
$\tilde{R}^2_{y.I\underline{PF}}$	.53@	.40@	.70@	.66@	.40@	.59@	.55@	.42@	.45@	.45@	.45@	.60@	.47@	.46@	.58@	.45@
$R^2_{y.\underline{PF}}-R^2_{y.\underline{F}}$	.03	.08	.03	.02	.01	.05	.03	.02	.09	.02	.05	.04	.04	.02	.02	.03
$\tilde{R}^2_{y.\underline{PF}}-\tilde{R}^2_{y.\underline{F}}$	.01	.05	.01	.00	.00	.02	.01	.00	.07	.00	.03	.02	.02	.00	.00	.02
$R^2_{y.I\underline{PF}}-R^2_{y.\underline{PF}}$	.16@	.03	.08@	.09*	.09@	.05@	.09@	.06@	.06	.01	.02	.01	.03	.01	.02	.02
$\tilde{R}^2_{y.I\underline{PF}}-\tilde{R}^2_{y.\underline{PF}}$	.16@	.02	.08@	.09*	.08@	.04@	.09@	.06@	.06	.00	.01	.01	.03	.00	.01	.01

1---Letters in ( ) refer to the variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; F=incomplete factor score composite variable; $\underline{F}$=complete factor score composite variable; I=incidental variable; $\underline{P}$=post-course motivation-related variable.
3---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$

Table 1-11A: Summary of Squared Multiple and Semi-Partial Correlations of Composite Ratings (IC) on General Rating (GI), Post-Course Motivation-Related Items (P), and Teacher Enthusiasm and Entertainment (I)

	(S1)[1] Skill				(R1) Rapport				(C1) Content				(X1) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.G}$[3]	.50	.52	.70	.67	.28	.51	.46	.45	.23	.24	.30	.38	.41	.20	.18	.33
$\tilde{R}^2_{y.G}$[4]	.49	.50	.69	.66	.26	.49	.44	.43	.21	.21	.28	.36	.40	.17	.15	.31
$R^2_{y.\underline{P}G}$	.56	.58	.70	.67	.33	.58	.48	.49	.52	.57	.58	.60	.52	.27	.39	.42
$\tilde{R}^2_{y.\underline{P}G}$	.54	.54	.68	.65	.29	.54	.44	.46	.49	.53	.55	.57	.49	.20	.34	.38
$R^2_{y.I\underline{P}G}$	.56	.60	.71	.70	.34	.59	.49	.49	.52	.57	.60	.61	.52	.29	.40	.43
$\tilde{R}^2_{y.I\underline{P}G}$	.53	.55	.68	.68	.30	.54	.44	.45	.49	.52	.56	.58	.49	.20	.34	.38
$R^2_{y.\underline{P}G}-R^2_{y.G}$	.06	.06	.00	.00	.05	.07	.02	.04	.29@	.33@	.28@	.22@	.11@	.07	.20@	.09@
$\tilde{R}^2_{y.\underline{P}G}-\tilde{R}^2_{y.G}$	.05	.04	.00	.00	.03	.05	.00	.03	.28@	.32@	.27@	.21@	.09@	.03	.19@	.07@
$R^2_{y.I\underline{P}G}-R^2_{y.\underline{P}G}$	.00	.02	.01	.03	.01	.01	.01	.00	.00	.00	.02	.01	.00	.02	.01	.01
$\tilde{R}^2_{y.I\underline{P}G}-\tilde{R}^2_{y.\underline{P}G}$	.00	.01	.00	.03	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

1---Items in ( ) are complete factor score rat ings.
2---Letter and number symbols refer to discipl ine (economics) and section.
3---Subscript notation: y=criterion variable; P=post-course motivation-related item; G=general item; I=incidental variable.
4---R^2 is the conventional notation for shrunk en R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-11B: Summary of Squared Multiple and Semi-Partial Correlations of Composite Ratings (IFS) on General Ratings (GI), Post-Course Motivation-Related Items (P), and Teacher Enthusiasm and Entertainment (I)

	(S2)[1] Skill				(R2) Rapport				(C2) Content				(X2) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.G}$[3]	.50	.55	.68	.68	.26	.49	.45	.41	.24	.25	.31	.40	.17	.19	.17	.31
$\tilde{R}^2_{y.G}$[4]	.49	.53	.67	.67	.24	.48	.43	.39	.22	.22	.29	.38	.15	.16	.14	.29
$R^2_{y.\underline{P}G}$	.55	.60	.68	.69	.30	.58	.46	.44	.52	.58	.58	.61	.27	.27	.39	.41
$\tilde{R}^2_{y.\underline{P}G}$	.53	.56	.66	.69	.26	.54	.42	.41	.49	.54	.55	.58	.23	.20	.34	.37
$R^2_{y.I\underline{P}G}$	.55	.62	.69	.70	.31	.59	.47	.45	.53	.58	.59	.62	.28	.29	.40	.42
$\tilde{R}^2_{y.I\underline{P}G}$	.52	.59	.66	.68	.27	.54	.42	.40	.50	.53	.55	.59	.23	.20	.34	.37
$R^2_{y.\underline{P}G}-R^2_{y.G}$	.05	.05	.00	.01	.04	.09	.01	.03	.28@	.33@	.27@	.21@	.10	.08@	.22@	.10@
$\tilde{R}^2_{y.\underline{P}G}-\tilde{R}^2_{y.G}$	.04	.03	.00	.00	.02	.06	.00	.02	.27@	.32@	.36@	.20@	.08@	.04@	.20@	.06*
$R^2_{y.I\underline{P}G}-R^2_{y.\underline{P}G}$	.00	.02	.01	.01	.01	.01	.01	.01	.01	.00	.01	.01	.01	.02	.01	.01
$\tilde{R}^2_{y.I\underline{P}G}-\tilde{R}^2_{y.\underline{P}G}$	.00	.01	.00	.01	.01	.00	.00	.00	.01	.00	.00	.01	.00	.00	.00	.00

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; P=post-course motivation-related item; G=general item; I=incidental variable.
4---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample size: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-11C: Summary of Squared Multiple and Semi-Partial Correlations of Composite Ratings (CFS) on General Ratings (GI), Post-Course Motivation-Related Items (P), and Teacher Enthusiasm and Entertainment (I)

	(S3)[1] Skill				(R3) Rapport				(C3) Content				(X3) Organization/ Structure			
	E1[2]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$R^2_{y.G}$[3]	.34	.48	.55	.52	.14@	.45	.27	.21	.11@	.08	.14	.21	.11@	.09	.08*	.21
$\tilde{R}^2_{y.G}$[4]	.33	.46	.54	.51	.12@	.44	.25	.19	.09@	.04	.11	.19	.09@	.05	.05	.19
$R^2_{y.PG}$	.38	.54	.57	.57	.18@	.53	.27	.27	.44	.44	.51	.51	.18	.15	.28	.34
$\tilde{R}^2_{y.PG}$	.36	.50	.55	.55	.14@	.49	.23	.22	.41	.38	.47	.48	.14@	.06	.22	.30
$R^2_{y.IPG}$	.38	.57	.59	.58	.19	.54	.30	.28	.45	.45	.53	.51	.20	.17	.30	.37
$\tilde{R}^2_{y.IPG}$	.35	.53	.56	.56	.15	.49	.25	.22	.43	.38	.48	.47	.15@	.06	.23@	.32@
$R^2_{y.PG}-R^2_{y.G}$	.04	.06	.02	.05	.04	.08	.00	.06	.33	.36	.37	.30	.07*	.06	.20@	.13@
$\tilde{R}^2_{y.PG}-\tilde{R}^2_{y.G}$	.03	.04	.01	.04	.02	.05	.00	.03	.32@	.34@	.36@	.29@	.05	.01	.17@	.11@
$R^2_{y.IPG}-R^2_{y.PG}$	.00	.03	.02	.01	.01	.01	.03	.01	.01	.01	.02	.00	.02	.02	.02	.03
$\tilde{R}^2_{y.IPG}-\tilde{R}^2_{y.PG}$	.00	.03	.01	.01	.01	.00	.02	.00	.02	.00	.01	.00	.01	.00	.01	.02

1---Items in ( ) are complete factor score ratings.
2---Letter and number symbols refer to discipline (economics) and section.
3---Subscript notation: y=criterion variable; G=general item; I=incidental variable; P=post-course motivation-related item;
4---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p.106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample size: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$.

Table 1-11D: Summary of Results from Hierarchical Regression An alysis for General (Composite) Ratings, Post-Course Motivation-Related Student Variables and Teacher Enthusiasm and Entertainment Items

	Skill				Rapport				Content				Organization/ Structure			
	E1[1]	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
$\tilde{R}^2_{y.I\underline{P}C}$[2]	.53@	.39@	.71@	.64@	.39@	.59@	.54@	.44@	.47@	.41@	.45@	.60@	.46@	.40@	.59@	.45@
$\tilde{R}^2_{y.I\underline{P}G}$[3]	.53@	.55@	.68@	.68@	.30@	.54@	.44@	.45@	.49@	.52@	.56@	.58@	.49@	.20@	.34@	.38@
$\tilde{R}^2_{y.C}$	.35@	.30@	.59@	.53@	.33@	.51@	.44@	.38@	.32@	.41@	.40@	.58@	.40@	.39@	.57@	.42@
$\tilde{R}^2_{y.G}$	.49@	.50@	.69@	.66@	.26@	.49@	.44@	.43@	.21@	.21@	.28@	.36@	.40@	.17@	.15@	.31@
$\tilde{R}^2_{y.\underline{P}C}-\tilde{R}^2_{y.C}$	.03	.06	.02	.01	.00	.04	.01	.00	.08	.00	.03	.02	.05	.00	.01	.03
$\tilde{R}^2_{y.\underline{P}G}-\tilde{R}^2_{y.G}$	.05	.04	.00	.00	.03	.05	.00	.03	.28@	.32@	.27@	.21@	.09@	.03@	.19@	.07@
$\tilde{R}^2_{y.I\underline{P}C}-\tilde{R}^2_{y.\underline{P}C}$	.15@	.03	.10@	.10@	.07@	.04*	.09@	.07@	.07@	.00	.02	.00	.01	.01	.01	.00
$\tilde{R}^2_{y.I\underline{P}G}-\tilde{R}^2_{y.\underline{P}G}$	.00	.01	.00	.03	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

1---Letter and number symbols refer to discipl ine (economics) and section.
2---Subscript notation: y=criterion variable; $\underline{P}$=post-course motivation-related item; I=incidental variable; G=general item; $\underline{C}$=composite item.
3---$\tilde{R}^2$ is the conventional notation for shrunk en R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample size: $N_{E1}=179$, $N_{E2}=98$, $N_{E3}=121$, $N_{E4}=142$

Table 1-12A: Intercorrelations between General and Composite Rating Items for Grouped Data; Economics Classes.

	GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	X2	S3	R3	C3	X3
(GS)																
(GR)	.62															
(GC)	.61	.48														
(GX)	.59	.47	.70													
(S1)	.75	.55	.67	.70												
(R1)	.58	.63	.42	.46	.61											
(C1)	.46	.44	.50	.47	.53	.46										
(X1)	.30	.25	.35	.37	.39	.31	.44									
(S2)	.75	.54	.66	.69	.98	.59	.51	.35								
(R2)	.57	.61	.41	.44	.58	.99	.44	.30	.57							
(C2)	.48	.45	.51	.48	.55	.47	.99	.45	.52	.45						
(X2)	.31	.25	.35	.37	.39	31	.44	.98	.35	.30	.34					
(S3)	.70	.46	.59	.63	.92	.47	.27	.17	.95	.44	.29	.18				
(R3)	.37	.50	.22	.24	.31	.91	.31	.12	.29	.93	.32	.11	.16			
(C3)	.29	.29	.37	.33	.34	.23	.94	.27	.32	.20	.94	.27	.09	.11		
(X3)	.15	.12	.22	.23	.21	.13	.30	.96	.15	.12	.31	.96	.00	.00	.15	

1---Letter symbols refer to the veriable meaning in the ori ginal questionnaire.
2---Letter and number symbols are complete factor score rat ings.
Sample sizes: N_E=546, $|r| \geq .12$, $P < .01$.

Table 1-12B: Intercorrelations between General and Composite Rating Items for Grouped Data; Statistics and German Classes.

		GS[1]	GR	GC	GX	S1[2]	R1	C1	X1	S2	R2	C2	S3	R3	C3	X3
(GS)		.79	.81	.74	.82	.58	.52	.63	.83	.54	.52	.61	.79	.24	.31	.48
(GR)	.66		.74	.69	.80	.69	.60	.72	.78	.65	.61	.70	.72	.37	.37	.62
(GC)	.41	.42		.78	.81	.71	.66	.69	.79	.67	.67	.68	.72	.41	.46	.58
(GX)	.64	.44	.59		.79	.72	.59	.65	.77	.68	.59	.66	.68	.45	.36	.55
(S1)	.68	.56	.65	.72		.75	.62	.72	.98	.73	.63	.72	.93	.41	.34	.58
(R1)	.52	.60	.42	.44	.59		.62	.66	.70	.99	.63	.72	.58	.87	.35	.58
(C1)	.37	.44	.511	.42	.61	.44		.61	.57	.60	.99	.61	.39	.40	.92	.50
(X1)	.17	.15	.34	.43	.37	.25	.37		.68	.64	.62	.98	.54	.38	.39	.96
(S2)	.67	.53	.63	.72	.98	.59	.59	.36		.67	.58	.68	.97	.33	.28	.51
(R2)	.51	.58	.42	.42	.60	.99	.42	.23	.59		.61	.65	.55	.89	.33	.56
(C2)	.38	.45	.52	.43	.63	.45	.99	.36	.61	.43		.62	.40	.41	.92	.51
(X2)	.12	.14	.33	.39	.30	.19	.31	.97	.29	.17	.29		.59	.38	.38	.97
(S3)	.63	.48	.55	.68	.93	.52	.38	.20	.96	.53	.41	.14		.19	.09	.40
(R3)	.37	.49	.25	.24	.35	.91	.26	.01	.33	.93	.28	-.03	.26		.21	.35
(C3)	.22	.29	.39	.26	.45	.24	.95	.20	.42	.21	.95	.14	.22	.09		.29
(X3)	.06	.10	.27	.31	.17	.09	.20	.94	.14	.06	.18	.97	.00	-.08	.04	

1---Letter symbols refer to the variable meaning in the original questionnaire.
2---Letter and number symbols are complete factor score ratings.
Sample sizes: N_S=110, $|r| \geq .26$, $P < .01$; N_G=67, $|r| \geq .31$, $P < .01$.

Table 1-13A: Intercorrelations of Background, Pre-Course and Post Course Motivation-Related Variables with Post-Course Motivation-Related Variables, General and Composite Ratings; Economics Classes

Post-Course Ratings		Pre-Course Ratings												Post-Course Ratings									
		(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(12)	(13)	(14)	(15)	(18)	(22)	(23)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Amount Learned	(3)[1]	.04	.01	-.04	.01	-.05	.01	.26@	.05	.07	.23	.10	-.01	.02	.04	.02	.17		.25	.10	.48	.29	.34	.39
Relevance	(4)	.12	.13	.01	.17@	.11	.10	.24@	.52	.11	.34	.20	.01	.20	.17	.13	.24	.25		.09	.45	.22	.18	.22
Workload	(5)	.06	.02	.15@	.13	.03	.05	.13	.12	.06	.02	.22	.01	.07	-.04	-.03	-.09	.09	.09		.04	.45	.02	-.06
Interest	(6)	.19@	.17@	-.02	.00	-.10	.12	.31	.16	.18	.52	.17	.04	.22	.23	.22	.31	.48	.45	.04		.30	.40	.46
Effort	(7)	.12	.06	.16	.13	.04	.11	.21	.14	.08	.15	.30	.02	.07	.03	.04	.08	.29	.22	.45	.30		.09	.07
Enthusiasm	(8)	.04	.02	.07	.02	-.07	-.01	.13	.05	.07	.20	-.02	.03	.06	.02	.03	.09	.33	.18	.02	.41	.09		.57
Entertainment	(9)	.11	.07	.01	-.02	-.15@	-.01	.21@	.07	.09	.27	.01	-.02	.09	.09	.04	.12	.39	.21	-.07	.46	.07	.57	
Skill	(GS)[2]	.03	-.02	-.01	.06	-.06	-.04	.26@	.02	.08	.26@	-.01	.07	.10	.10	.07	.10	.50	.17	-.02	.48	.13	.64	.72
Rapport	(GR)	.07	.03	.00	.05	-.07	-.06	.18	.00	.11	.19	.05	.08	.02	.03	.04	.05	.33	.17	.00	.40	.11	.56	.52
Content	(GC)	.12	.09	-.06	-.07	-.19@	.03	.20	.06	.09	.26	.05	.04	.08	.02	.00	.10	.50	.20	-.04	.48	.09	.51	.51
Organization/St	(GX)	.05	.02	-.04	-.03	-.14	.07	.23@	.05	.12	.30	.07	.02	.11	.05	.04	.17	.46	.25	.00	.49	.11	.50	.51
Skill	(S1)[3]	.09	.05	.01	-.03	-.12	.05	.26@	.05	.07	.28	.01	.02	.07	.08	.03	.18	.48	.21	-.09	.49	.09	.60	.63
Rapport	(R1)	.08	.06	.03	.04	-.05	.01	.25@	.07	.12	.27	.10	-.01	.11	.04	.04	.10	.35	.19	.00	.41	.12	.50	.52
Content	(C1)	.11	.12	.01	.01	-.14	.08	.31@	.22	.19	.45	.15	.05	.20	.09	.09	.16	.47	.50	.02	.67	.28	.37	.43
Organization/St	(X1)	-.11	-.06	.02	.03	-.01	-.02	.16	.06	.06	.21	.05	.05	.08	-.08	-.10	.17	.41	.41	-.07	.28	.09	.27	.17

1---Numbers in ( ) refer to the variable numbers in the orig inal questionnaire.
2---Letter symbols refer to the variable meaning in the orig inal questionnaire.
3---Letter and number symbols are complete factor score rati ings.
@ = $p \leq .01$
Correlations for pre- by post-course items based upon N=343, $r \geq .15$, $p < .01$.
Correlations for post- by pre-course items based upon N=546, $r \geq .12$, $p < .01$.

Table 1-13B: Intercorrelations of Background, Pre-Course and Post-Course Motivation-Related Variables with Post-Course Motivation-Related Variables, General and Composite Ratings; Statistics Classes

Post-Course Ratings	Pre-Course Ratings																					
		(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(12)	(13)	(14)	(15)	(18)	(22)	(23)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Amount Learned	(3)	-.17	-.28	-.10	.14	.25	-.02	.42	.34	.20	.31	.19	-.17	.11	.00	-.03	.36		.47	.03	.59	.29	.60	.59
Relevance	(4)	-.34	-.26	.16	.30	.35	-.07	.56	.59	.30	.55	.11	.05	.12	.15	.07	.17	.47		.02	.63	.30	.38	.37
Workload	(5)	-.12	.07	-.09	.23	.12	-.19	.31	-.11	-.03	.19	.33	.26	-.27	-.06	-.04	-.05	.03	.03		.04	.36	.01	-.09
Interest	(6)	-.21	-.06	.10	.22	.31	.11	.57	.49	.28	.59	.07	-.10	-.21	.24	.05	.41	.59	.63	.04		.29	.40	.48
Effort	(7)	-.17	-.16	.10	.16	.39	-.20	.14	.11	-.08	.09	.32	.01	-.11	.01	-.20	.21	.29	.30	.36	.29		.22	.11
Enthusiasm	(8)	-.21	.00	.12	.07	-.02	-.03	.37	.32	.22	.20	.13	.08	-.05	-.04	-.04	.28	.60	.38	.01	.40	.22		.70
Entertainment	(9)	.10	-.03	-.02	.18	.16	.21	.26	.24	.21	.21	-.08	.11	.08	.26	.25	.36	.59	.37	-.09	.48	.11	.70	
Skill	(GS)	-.15	-.02	.06	.14	.03	-.13	.38	.29	.15	.20	.09	-.02	.06	.01	.07	.34	.62	.41	.02	.47	.18	.78	.77
Rapport	(GR)	-.21	-.09	.02	.22	.00	-.09	.40	.33	.08	.32	-.06	.05	-.16	-.05	.02	.28	.54	.43	.02	.54	.18	.66	.70
Content	(GC)	-.21	-.14	-.02	.01	.04	-.04	.49	.38	.25	.35	.08	-.07	-.09	.12	.15	.42	.70	.46	.01	.60	.22	.72	.71
Organization/St	(GX)	-.20	-.12	.05	-.06	.00	.10	.38	.30	.22	.30	.07	-.10	.02	.12	.04	.39	.59	.41	-.-5	.49	.18	.67	.60
Skill	(S1)	-.30	-.22	-.01	-.11	.03	-.01	.40	.23	.13	.25	.03	-.14	-.02	-.05	-.16	.41	.62	.51	.02	.57	.24	.73	.70
Rapport	(R1)	-.18	-.18	-.13	.10	.07	-.07	.38	.21	.17	.19	-.03	-.11	-.08	.05	-.09	.37	.55	.38	.02	.54	.22	.55	.54
Content	(C1)	-.17	-.15	.15	.21	.28	-.01	.47	.45	.30	.47	.25	-.11	-.19	.29	.17	.43	.54	.71	.10	.76	.34	.35	.50
Organization/St	(X1)	-.17	-.13	.07	.20	.01	.18	.34	.28	-.05	.26	-.05	-.09	-.17	.13	.00	.34	.47	.41	-.03	.53	.17	.54	.58

1---Numbers in ( ) refer to the variable numbers in the original questionnaire.
2---Letter symbols refer to the variable meaning in the original questionnaire.
3---Letter and number symbols are complete factor score ratings.
Correlations for pre- by post-course items based upon N=56, $r \geq .34$, $P < .01$.
Correlations for post- by pre-course items based upon N=110, $r \geq .26$, $P < .01$.

Table 1-13C: Intercorrelations of Background, Pre-Course and Post-Course Motivation-Related Variables with Post-Course Motivation-Related Variables, General and Composite Ratings; German Classes

Post-Course Ratings		Pre-Course Ratings												Post-Course Ratings									
		(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(12)	(13)	(14)	(15)	(18)	(22)	(23)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Amount Learned	(3)	.05	.04	-.15	.10	.01	.03	.14	.21	.02	.10	.04	.10	.00	.05	.08	.31		.25	-.09	.36	-.05	.32	.51
Relevance	(4)	.01	.07	-.14	.07	-.14	.26	.25	.79	.03	.15	-.14	.04	.08	.14	.14	.22	.25		.02	.33	-.07	.01	.18
Workload	(5)	-.02	-.03	.00	-.02	.19	-.24	-.03	-.21	.15	-.25	.37	.01	-.09	-.24	-.39	-.23	-.09	.01		-.06	.50	.30	.16
Interest	(6)	.09	.09	.01	.04	-.13	.21	.32	.27	.10	.64	.05	-.06	.00	.19	.20	.30	.36	.33	-.06		.16	.31	.36
Effort	(7)	-.06	-.09	.13	.09	.26	.00	.21	-.13	.16	.14	.56	-.19	.01	-.12	-.30	-.25	-.05	-.07	..50	.16		.12	.10
Enthusiasm	(8)	.06	-.03	.12	.00	.18	-.28	.17	-.18	.25	.13	.27	.25	.05	-.37	-.24	-.08	.32	.01	.30	.31	.12		.43
Entertainment	(9)	-.03	-.07	-.01	.13	-.05	-.23	.11	.19	.19	.00	-.10	-.04	-.22	-.13	-.30	.06	.51	.18	.16	.36	.10	.43	
Skill	(GS)	.03	-.02	.13	.04	-.14	-.04	.27	.27	-.04	.15	-.01	.12	.02	-.14	-.21	.16	.44	.22	.10	.29	-.03	.56	.74
Rapport	(GR)	.11	.02	.07	-.07	.15	-.14	.32	-.01	.08	.04	.24	.21	-.01	-.18	-.23	.04	.33	.19	.37	.41	.17	.56	.58
Content	(GC)	.20	.19	-.26	-.12	-.20	-.26	-.05	.00	.09	.02	.00	.09	.31	.04	.01	.11	.47	.14	-.01	.29	.00	.29	.43
Organization/St	(GX)	.10	.15	-.14	-.16	-.20	-.03	.04	.13	-.01	.06	-.09	-.08	.18	-.07	-.08	.19	.37	.15	.12	.27	.11	.31	.58
Skill	(S1)	.03	.02	-.08	-.07	-.16	.03	.03	.29	-.02	.18	-.04	-.01	.17	-.05	-.09	.28	.51	.35	.10	.39	-.02	.35	.58
Rapport	(R1)	.03	-.01	-.05	-.04	.06	-.08	.02	-.1-	.00	.04	.20	.05	.01	-.14	-.06	.14	.46	.09	.33	.33	.12	.64	.47
Content	(C1)	-.02	.04	-.13	-.12	-.21	.07	.18	.49	.01	.26	-.09	.01	.04	.13	.29	.24	.45	.56	.01	.71	.01	.25	.38
Organization/St	(X1)	.12	.14	.11	-.22	-.20	-.12	.00	.11	.03	-.09	.18	-.09	.10	.17	-.13	.11	.13	.12	.08	.16	.14	.12	.29

1---Numbers in ( ) refer to the variable numbers in the original questionnaire.
2---Letter symbols refer to the variable meaning in the original questionnaire.
Correlations for pre- by post-course items based upon N=46, $r \geq .37$, $P < .01$.
Correlations for post- by pre-course items based upon N=67, $r \geq .31$, $P < .01$.

Table 1-14A: Squared Multiple and Semi-Partial Correlations of General Ratings (GI) Regressed on Background (B), Pre-Course (P) and Post-Course (P) Motivation-Related Characteristics for Grouped Data; Statistics (S), German (G), Economics (E)

	(GS)[1] Skill			(GR) Rapport			(GC) Content			(GX) Organization/ Structure		
	S	G	E	S	G	E	S	G	E	S	G	E
$R^2_{y.B}$[2]	.05	.09	.02	.10	.07	.03	.05	.13	.05@	.06	.08	.03
$\tilde{R}^2_{y.B}$[3]	.00	.00	.01	.00	.00	.02	.00	.02	.04	.00	.00	.02
$R^2_{y.PB}$	.20	.18	.13@	.28	.20	.08@	.32	.15	.12@	.24	.10	.12
$\tilde{R}^2_{y.PB}$	.00	.00	.10	.10	.00	.00	.15	.00	.09	.05	.00	.09
$R^2_{y.\underline{P}PB}$	.50*	.46	.37@	.50*	.57*	.21@	.63@	.52	.36@	.52	.37	.32
$\tilde{R}^2_{y.\underline{P}PB}$	.29	.17	.34@	.29	.34@	.17@	.46	.26	.33@	.31	.03	.29
$R^2_{y.PB}-R^2_{y.B}$	.15	.09	.11@	.18	.13	.05@	.27*	.02	.07@	.18	.02	.09
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.00	.00	.09	.10	.00	.00	.15	.00	.05	.05	.00	.07
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.30	.28	.24@	.22	.37@	.13@	.31@	.37@	.24@	.28	.27	.20
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.29*	.17@	.24@	.19	.34@	.17@	.31@	.26	.24@	.26	.03	.20

1---Letter symbols refer to variable meaning in the original questionnaire.
2---Subscript notation: y=criterion variable; B= background variable; P=pre-course motivation-related items; $\underline{P}$=post-course item.
3---$\tilde{R}^2$ is the conventional notation for shrunker R^2 (see Cohen and Cohen, 1975, p. 106-107).
* = $P \leq .05$
@ = $P \leq .01$
Sample sizes: N_S=51, N_G=45, N_E=330.

Table 1-14B: Squared Multiple and Semi-Partial Correlations of Composite Ratings (IC) Regressed on Background (B), Pre-Course (P) and Post-Course (P) Motivation-Related Characteristics for Grouped Data; Statistics (S),German (G), Economics (E)

	(S1)[1] Skill			(R1) Rapport			(C1) Content			(X1) Organization/ Structure		
	S	G	E	S	G	E	S	G	E	S	G	E
$R^2_{y.B}$	.12	.03	.03	.07	.01	.02	.16	.07	.04*	.08	.12	.02
$\tilde{R}^2_{y.B}$[2]	.02	.00	.02	.00	.00	.01	.07	.00	.03	.00	.00	.01
$R^2_{y.PB}$	.29	.13	.10@	.23	.06	.10@	.48@	.33@	.23@	.24	.19	.07@
$\tilde{R}^2_{y.PB}$	.12	.00	.07	.14	.00	.07@	.35	.13	.21@	.05	.00	.03
$R^2_{y.\underline{P}PB}$	.80@	.45	.37@	.52*	.45	.21@	.79@	.78@	.56@	.50*	.00	.23@
$\tilde{R}^2_{y.\underline{P}PB}$	.71@	.11	.34@	.31	.11	.17@	.70@	.66	.55@	.29	.00	.19@
$R^2_{y.PB}-R^2_{y.B}$	.17	.10	.07@	.16	.05	.08@	.32*	.26	.19@	.16	.07	.05*
$\tilde{R}^2_{y.PB}-\tilde{R}^2_{y.B}$	.10	.00	.05	.14	.00	.06@	.28*	.13	.18@	.05	.00	.02
$R^2_{y.\underline{P}PB}-R^2_{y.PB}$	.57@	.32*	.27@	.29@	.39@	.11@	.31@	.45	.33@	.26	.09	.16@
$\tilde{R}^2_{y.\underline{P}PB}-\tilde{R}^2_{y.PB}$	.57@	.11@	.27@	.17	.11	.10@	.31@	.45@	.35@	.26	.09	.16@

1---Letter and number symbols are complete factor score ratings.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (See Cohen and Cohen, 1975, p. 106-107).
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: N_S=51, N_G=45, N_E=330.

Table 1-15A: Squared Multiple and Semi-Partial Correlations from Regression of General Ratings (GI) on Composite Ratings (IC), Post-Course Motivation-Related Items (P), and Teacher Entertainment and Enthusiasm (I) for Grouped Data: Statistics (S), German (G), Economics (E)

	(GS)[1] Skill			(GR) Rapport			(GC) Content			(GX) Organization/ Structure		
	S	G	E	S	G	E	S	G	E	S	G	E
$R^2_{y.C}$	.69@	.50@	.59@	.70@	.67@	.45@	.71@	.45@	.48@	.67@	.56@	.50@
$\tilde{R}^2_{y.C}$[2]	.68@	.47@	.59@	.69@	.65@	.45@	.70@	.41@	.48@	.66@	.53@	.50@
$R^2_{y.\underline{P}C}$	.72@	.51@	.61@	.70@	.72@	.46@	.76@	.52@	.52@	.69@	.57@	.53@
$\tilde{R}^2_{y.\underline{P}C}$	.69@	.43@	.60@	.67@	.67@	.45@	.74@	.44@	.51@	.66@	.50@	.52@
$R^2_{y.I\underline{P}C}$	.80@	.73@	.70@	.73@	.77@	.51@	.79@	.52@	.53@	.71@	.61@	.53@
$\tilde{R}^2_{y.I\underline{P}C}$	.78@	.67@	.69@	.70@	.72@	.50@	.77@	.42@	.52@	.68@	.53@	.52@
$R^2_{y.\underline{P}C}-R^2_{y.C}$	.03	.01	.02	.00	.05	.01	.05	.07	.04	.02	.01	.03
$\tilde{R}^2_{y.\underline{P}C}-\tilde{R}^2_{y.C}$	.01	.00	.01	.00	.02	.00	.04	.03	.03	.00	.00	.02
$R^2_{y.I\underline{P}C}-R^2_{y.\underline{P}C}$	.08@	.22@	.09@	.03@	.05@	.05@	.03*	.00	.01	.02	.04	.00
$\tilde{R}^2_{y.I\underline{P}C}-\tilde{R}^2_{y.\underline{P}C}$	.08@	.22@	.09@	.03*	.05@	.05@	.03*	.00	.01	.02	.03	.00

1---Letter symbols refer to variable meaning in the original questionnaire.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen and Cohen, 1975, p. 106-107).
@ = $p \leq .01$
* = $p \leq .05$
Sample sizes: N_S=104, N_G=63, N_E=546.

Table 1-15B: Squared Multiple and Semi-Partial Correlations from Regression of Composite Ratings (IC) on General Ratings (GI), Post-Course Motivation-Related Items (P), and Teacher Entertainment and Enthusiasm (I) for Grouped Data: Statistics (S), German (G), Economics (E)

	(S1)[1] Skill			(R1) Rapport			(C1) Content			(X1) Organization/ Structure		
	S	G	E	S	G	E	S	G	E	S	G	E
$R^2_{y.G}$	.80@	.68@	.67@	.64@	.42@	.47@	.48@	.33@	.32@	.59@	.21@	.16@
$\tilde{R}^2_{y.G}$[2]	.79@	.66@	.67@	.63@	.38@	.47@	.46@	.28@	.32@	.57@	.16*	.15@
$R^2_{y.\underline{P}G}$	.81@	.73@	.68@	.65@	.52@	.47@	.74@	.73@	.57@	.60@	.23	.23@
$\tilde{R}^2_{y.\underline{P}G}$	.78@	.68@	.68@	.62@	.44@	.46@	.72@	.68@	.56@	.56@	.10	.22@
$R^2_{y.I\underline{P}G}$	.81@	.73@	.69@	.65@	.58@	.48@	.79@	.73@	.57@	.61@	.26@	.25@
$\tilde{R}^2_{y.I\underline{P}G}$	.79@	.67@	.68@	.61@	.49@	.47@	.77@	.67@	.56@	.56	.10	.24@
$R^2_{y.\underline{P}G}-R^2_{y.G}$	.01	.05	.01	.01	.10	.00	.26@	.40@	.25@	.01	.02	.07@
$\tilde{R}^2_{y.\underline{P}G}-\tilde{R}^2_{y.G}$	.00	.02	.01	.00	.06	.00	.26@	.40@	.24@	.00	.00	.07@
$R^2_{y.I\underline{P}G}-R^2_{y.\underline{P}G}$	.00	.00	.01	.00	.06	.01	.05	.00	.00	.01	.03	.02
$\tilde{R}^2_{y.I\underline{P}G}-\tilde{R}^2_{y.\underline{P}G}$	.00	.00	.00	.00	.05	.01	.05	.00	.00	.00	.00	.02

1---Letter and number symbols are complete factor score ratings.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2.
@ = $p \leq .01$
* = $p \leq .05$
Sample sizes: N_S=104, N_G=63, N_E=546.

Table 1-16: Squared Multiple and Semi-Partial Correlations from Regression of Pre-Course Motivation-Related Characteristics on Background (B) and Related Pre-Course Items (1-3) for Grouped Data: Statistics (S), German (G), Economics (E)

	(9)[1] Amount Learned			(10) Relevance			(12) Workload			(13) Interest			(14) Effort		
	S	G	E	S	G	E	S	G	E	S	G	E	S	G	E
$R^2_{y.B}$	.05	.12	.01	.03	.11	.07@	.04	.08	.03*	.05	.07	.05@	.05	.11	.06@
$\tilde{R}^2_{y.B}$[2]	.00	.01	.00	.00	.00	.06@	.00	.00	.02	.00	.00	.04	.00	.00	.05
$R^2_{y.1B}$[3]	.06	.00	.06@	.00	.00	.16@	.00	.14	.08@	.01	.09	.19@	.00	.12	.11@
$\tilde{R}^2_{y.1B}$	.00	.00	.04*	.00	.00	.14@	.00	.05	.06@	.00	.00	.18@	.00	.03	.09@
$R^2_{y.321B}$	.01	.00	.07@	.00	.12	.16@	.05	.14	.08@	.00	.09	.21@	.01	.14	.13@
$\tilde{R}^2_{y.321B}$	.00	.00	.05*	.00	.00	.14@	.00	.01	.06*	.00	.00	.19@	.00	.01	.11@
$R^2_{y.1B}-R^2_{y.B}$	.01	.00	.05*	.00	.00	.09@	.00	.06	.05	.00	.02	.14@	.00	.01	.05
$\tilde{R}^2_{y.1B}-\tilde{R}^2_{y.B}$	.00	.00	.04*	.00	.00	.08@	.00	.05	.04	.00	.00	.14@	.00	.03	.04
$R^2_{y.321B}-R^2_{y.1B}$	.00	.00	.01	.00	.12	.00	.05	.00	.00	.00	.00	.02	.01	.02	.02
$\tilde{R}^2_{y.321B}-\tilde{R}^2_{y.1B}$	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.02

1---Numbers in ( ) refer to the variable numbers in the original questionnaire.
2---$\tilde{R}^2$ is the conventional notation for shrunken R^2 (see Cohen & Cohen, 1975, p. 106-107).
3---Subscript notation: 1=liking for social sciences; 2=average grade for prior course; 3=overall teacher/course rating for prior course.
@ = $P \leq .01$
* = $P \leq .05$
Sample sizes: N_S=51, N_G=45, N_E=330

ADDENDUM

The Effect of Course Content Upon the Factor Structure of Student Ratings: A Maximum Likelihood Analysis

Facultymembers typically object strongly to using a rating instrument to evaluate courses in different disciplines when the rating data becomesa means to compare or select a "best" teacher. That is, faculty members argue that the course content can dramatically affect ratings. This, in fact, may be the case for general rating items used to assess the "overall quality" of a teacher or course. For example, the content of a calculus course as opposed to an abnormal psychology course might cause studentsto feelthat they had learned more and, as a result, rate calculus instructors more highly. Perhaps, and probably more likely, calculus would be perceived as more abstract and less relevant, and therefore, the mathematics faculty will receive poorer student evaluations than psychology professors.

Rayder (1968) reports having faculty in the School of Education at Colorado State College evaluated by over four thousand undergaduates in eighty-seven courses. The basic idea of the research was to ascertain the extent of the influence that background variables, like student sex, grade point average, etc. had on ratings. The effect of these variables was consistent with the literature already reviewed in this proposal: these variables had little or no influence on ratings. However, one other aspect

of this study is important here. Rayder also attempted to use the teacher's "subject area" as a predictor variable in a regression analysis to measure its effect on three general ratings. Results indicate that subject area (by department in the school of education) accounted for less than ten percent of the variability in each of the general ratings.

Unfortunately, Rayder does not provide enough information to determine exactly how "subject areas" were classified (or divided) or the significance levels of the predictor variable. Presumably, the "subject area" variable was based upon the means of the overall ratings for all courses in a particular subject area.

Even though Rayder's evidence is inconclusive, it is important to note the suggestion of an effect of content area upon general teacher ratings. That is, student ratings may not be generalizable over content areas. Centra (1973) also supports this hypothesis by speculating about differences between students' perceptions of "hard" science courses and humanities although he provides no empirical evidence for his intutitions.

The purpose of the research described in this section will be to determine the effect of content area (between disciplines) upon the generalizability of the factor structure derived from the specific items of a rating

instrument. The null hypothesis will be, of course, that no differences in factor structure should occur due to content. However, intuitively, one would expect large differences in factor structures when comparing ratings of social sciences course (i.e., economics) with mathematics courses (i.e., calculus), or language courses (i.e., German).

Method

Subjects. Courses in four section of economics, six sections of statistics and five sections of German were rated with a common questionnaire at the University of Massachusetts. The total sample sizes by "subject area" are 962, 204 and 100, respectively. The student populations in each of the courses are composed of freshmen, sophomores, juniors and seniors taking these courses as a requirement or elective. The teachers are all full-time University of Massachusetts faculty members meeting three class hours per week with their students. In both economics and German courses, teachers are giving common final exams. All sections of German have, in addition, common texts, and syllabi. Portions of the economics syllabi are also common (covering the same substantive topic areas) utilizing identical texts and final exams with some common test questions. Alternatively, the mathematics sections are using a common

text but not following common syllabi, nor are the faculty members collaborating on a common final exam. However, in this case, the nature of the course material will allow selection of similar problem types and, therefore, provide a rough index of achievement for each section.

Design and quantitative methodology. Rayder (1968) attempted to determine the effect of course content area and/or discipline upon teacher rating outcomes. His data did not prove to be very useful, unfortunately, in providing an answer to this question. Even though this current research will examine a similar question (e.g., the effect of discipline upon the factor structure of rating instruments), the experimental groups (German, economics, and mathematics courses) and data analysis techniques will be quite different.

The subject areas of economics, German and statistics were selected under the assumption that effective teaching methods within each discipline might be quite different. Consequently, the question of whether or not the five basic dimensions (rapport, skill, organization, difficulty, and general course attitude) described in earlier sections of this paper will remain invariant over the different disciplines will be considered. To be a bit more specific, maximum likelihood factor analysis techniques (Joreskog, 1971, a, b; Mulaik, 1972) will be used to demonstrate the

invariance of the factor pattern matrix, the factor score variance-covariance matrix, and also the variance of the unique scores in the typical factor analysis model.

If content area has no effect on students' assessments of college teachers' teaching abilities, the factor structures between groups should be identical. The analytic techniques necessary to resolve this question have only recently been developed (Joreskog, 1966) and the actual computer programs perfected by Van Thillo and Joreskog (1970). These methods have not been widely used in social sciences research. Therefore, a thorough discussion of these methods will undoubtedly benefit those reading this dissertation. The discussion that follows in the next few pages will not include proofs. Complete references will be provided to allow the interested reader to return to the original articles whenever he/she is motivated to do so.

Factor analytic techniques can be described along a continuum (Mulaik, 1972) with exploratory methods at one end and confirmatory methods at the other. Exploratory factor analytic techniques are the methods most widely used by contemporary researchers. More often than not, these researchers use exploratory methods to reduce a large set of variables into a smaller number of easily understood underlying dimensions (factors). These

dimensions are then attributed with a certain amount of "meaningfulness" within the context of the experiment and the meanings of the original set of variables.

Confirmatory factor analysis, like exploratory factor analysis, is a label descriptive of a number of different methods which may be used to demonstrate that a set of a priori dimensions are replicable under different experimental conditions, especially, different groups of subjects. Some confirmatory analytic techniques allow researchers to "test" hypotheses about the factor structure of observed data, perhaps collected from several different populations. Excellent examples of this type of methodology are presented by Joreskog (1971, a, b) and labelled as maximum likelihood factor analysis. Other methods also seek to determine the similarity of the observed factor structure to some a priori structure, but do not use statistical tests of significance. These methods are called Procrustean techniques and the goodness-of-fit criteria are typically functions describing the least squares fit between the observed and hypothetical structures (Evans, 1971; Mulaik, 1972).

In the paragraphs which follow, Procrustean techniques will be described briefly as a vehicle by which to introduce and point out the major methodological

advantages one gains by using the more powerful maximum likelihood procedures. The discussion will describe the two types of techniques in the order in which they have been introduced, but first, a brief digression to present the general factor analysis model and define terminology.

The general model for factor analysis (Mulaik, 1972; Joreskog, 1971) can be represented as:

$$y = Fx + z. \qquad (1\text{-}1)$$

The (p x 1) random vector, y, of observed variables can be equated with a factor pattern matrix, F, of order (p x k), a random vector of factor scores, x, of order (k x 1), and a random vector, z, of unique scores of order (p x 1). It is typically assumed that $E(y) = E(x) = E(z) = 0$, $E(yy') = \Sigma$, $E(xx') = \Phi$, and $E(zz') = \Psi$. These assumptions usually imply that the variance-covariance matrix, Σ, has the following form:

$$\Sigma = F \Phi F' + \Psi. \qquad (1\text{-}2)$$

It is usually assumed that Φ is the identity matrix when the factors are to be orthogonal to one another and Ψ is a diagonal matrix of unique score variances. The equation (1-2) is the general structural model for both Procrustean and maximum likelihood factor analytic methods. As we will see, the differences between these methods is entirely due to the measures of "goodness-of-fit" between the

structure of the observed data and the a priori or hypothetical structure.

In orthogonal Procrustean factor analysis, as one example, the observed factor pattern matrix F is rotated by an orthogonal transformation matrix, T, of order (k x k), under the constraint that (TT' - I) = 0, to the target matrix, F* (Mulaik, 1972; Evan, 1971):

$$FT = F^*. \qquad (1\text{-}3)$$

The target matrix, F*, has the desired a priori structure to which the observed data is being transformed. One popular goodness-of-fit criterion is "that the sum of squares of the matrix E = (F*-FT) to be a minimum under the restraint that T is an orthogonal matrix" (Mulaik, 1972, p. 294). This is typically done by minimizing $f = \text{Tr}\,(EE') = \text{Tr}\,[(F^* - FT)(F^*-FT)']$. The problem, then, becomes one of finding the transformation matrix, T, which satisfies the conditions described above. The matrix can be obtained by a complex process involving Lagrangian multiples or the use of algorithms which approximate the Lagrange process. Complete details and proof of this procedure are provided on pages 295 through 299 of Mulaik (1972).

The previous discussion of the orthogonal Procrustean factor analytic technique does not adequately represent the many different types of solutions available using

this methodology, but it does provide the reader with a simplistic understanding of the process for comparative purposes. Prior to proceeding to the maximum likelihood techniques, several other points should be noted: a) the matrix F* can be completely specified, partially specified (confirmatory factor analysis) or completely unspecified (exploratory factor analysis); b) the major problem associated with orthogonal Procrustean solutions and other types of solutions (e.g., oblique solutions) is the criteria for goodness-of-fit: they do not allow for statistical judgment of the similarity of the factor structures. However, Joreskog (1966) has avoided this problem by applying the likelihood ratio technique to the problem of fitting observed data to a specified structure. This method does allow for statistical decisions about the similarity between the observed and hypothetical factor structures.

One might want to argue that the method of maximum likelihood factor analysis is in effect a special case of Procrustean analysis, but with a statistical criterion for goodness-of-fit. While it is true that the general model, (1-2) describes the structure of the variance-covariance matrix of observed scores in both techniques, there is no attempt to rotate the observed factor pattern to a specified structure in maximum likelihood factor

analysis as there is in the Procrustean methods. What occurs in the maximum likelihood method is the following: the observed data are used to produce the maximum likelihood estimates of F, Φ, and Ψ under the null and alternative hypotheses which are then used in the appropriate maximum likelihood functions as the numerator and denominator in the log likelihood ratio criterion, respectively. This ratio, under center conditions, approximates a large sample chi-square statistic. Before describing these techniques in detail, a brief but necessary digression.

The general model (1-2) will be modified to deal with several experimental groups (economics, German and mathematics classes) used in the current research. In particular, the model (1-2) will now read:

$$\Sigma_g = F_g \Phi_g F_g' + \Psi_g \qquad (1\text{-}4)$$

where g = 1 to m(m =3) for the number of experimental groups. The initial question of interest will be whether or not F is invariant over all experimental groups of subjects and that F is composed of five factors or dimensions (Skill, Rapport, Organization, Difficulty, and General Course Attitude) as discussed elsewhere in this dissertation.

The following assumptions will be required in the uses of maximum likelihood factor analytic techniques for several populations:

1. The several populations must be clearly defined and independent of one another.
2. The observed variables have a multinormal distribution in each population. [This is necessary to insure that each observed variance-covariance matrix, S_g, will have a Wishart Distribution (Joreskog, 1971)].
3. Each S_g has a Wishart Distribution (then the maximum likelihood methodology yields a large sample chi-square statistic as a measure of goodness-of-fit for the overall analysis).
4. The usual assumptions of the factor analysis model (1-2) hold in each population.

In addition, a special case is assumed to hold for this research. That is, the same variables are being measured in each of the three experimental groups and the entire factor pattern is invariant over those groups. This means that the same variables load on the same factors in each group and the loading of each variable remains approximately identical in magnitude in the factor pattern of each experimental group.

The analytic procedure. The first hypothesis of importance will be the hypothesis of equality of the variance-covariance matrices of the observed variables, i.e.,

$$H_{\Sigma}: \Sigma_1 = \Sigma_2 = \Sigma_3. \qquad (1\text{-}6)$$

If it can be shown that the variance-covariance matrices are identical, then it becomes an easy process to estimate the number of factors in each experimental group when the covariance matrices, Σ_g, all have the same structure specified in (1-2). The test statistic for this hypothesis is approximated by:

$$M = N \cdot \log|S| - \sum_{g=1}^{3} (n_g \cdot \log|S_g|) \quad (1\text{-}7)$$

where $S = \frac{1}{n} \sum_{g=1}^{3} n_g S_g$, the pooled variance-covariance matrix and $N = \sum_{g=1}^{3} n_g$, the total number of subjects in all three groups. It should be noted that M is distributed approximately as a chi-square with degrees of freedom:

$$d = \tfrac{1}{2}(n-1)p(p+1). \quad (1\text{-}8)$$

If M is non-significant (the variance-covariance matrix is invariant over the experimental population), it becomes a relatively simple procedure to determine the appropriate number of factors in F which must be invariant over groups (see Joreskog, 1969). However, should M produce a significant result, then H is untenable and we would want to proceed with additional analyses to determine the similarities as well as the dissimilarities in the factor structures (e.g., F_g, Φ_g, or Ψ_g).

The hypothesis of equality of number of common factors, i.e., H_k: $k_1 = k_2 = k_3$. The test of this hypothesis is achieved by an unrestricted factor analysis with k^2 elements fixed in F_g and/or Φ_g. A solution is unrestricted when

the number of fixed elements is less than or equal to k^2, the square of the number of factors (or columns of the (k x k) transformation matrix, T). If our equality of number of factors is to be based upon an orthogonal solution, then set and fix $\frac{1}{2}k(k-1)$ zeros in F_g. Alternatively, if the solution is to be oblique, then fix diagonal $\Phi_g = I$ and $k(k-1)$ zeros in F_g. Each analysis yields a chi-quare with $d_k = \frac{1}{2}[(p-k)^2 - (p+k)]$ degrees of freedom. The overall number of degrees of freedom is possible because the groups are independent, so the log likelihood ratio statistics can be added (Joreskog, 1966, 1969, 1971 a). The ratio itself for each group, stated as:

$$L_{\omega g} = -\tfrac{1}{2}n_g[\log|\Sigma_g| + \mathrm{Tr}(S_g\Sigma_g^{-1})] \qquad (1\text{-}9)$$

represents the maximum likelihood of observing the matrix Σ in the set of all matrices of the form (1-2) and let

$$L_{\Omega g} = -\tfrac{1}{2}n_g(\log|S| + p) \qquad (1\text{-}10)$$

represent the maximum likelihood for observing the matrix Σ in the set of all positive definite matrices of order (p x p). The log likelihood ratio is then $L = (L_{\omega g}/L_{\Omega g})$. However, rather than maximize the log likelihood ratio for each group, Joreskog (1966, 1967, 1969, 1971 a, b) argues that it is easier to minimize:

$$F = \tfrac{1}{2}\sum_{g=1}^{3} n_g[\log|\Sigma_g| + \mathrm{Tr}(S_g\Sigma_g^{-1}) - \log|S_g| - p_g]. \qquad (1\text{-}11)$$

The statistic for the overall hypothesis, H_k, for all

groups is just two (2) times minimum F (a proof of the relationship between the likelihood ratio, L, and equation (1-11) is provided in Joreskog, 1967).

If the hypothesis of a common number of factors is tenable, then we proceed to test the hypothesis of an invariant factor pattern, i.e., H_F: $F_1 = F_2 = F_3$. This common pattern may be completely unspecified or specified to have [(k-1) or more] zeros in certain positions plus one fixed non-zero element in each column of F. This means there will be $g \geq k^2$ fixed elements in the specified case. The chi-square is obtained by estimating $F, \Phi_1, \Phi_2, \Phi_3, \Psi_1, \Psi_2$, and Ψ_3 from S_1, S_2, and S_3 (the observed variance-covariance matrices) simultaneously. This yields a minimum value of F, such that twice F is a chi-square; χ_F^2 with degrees of freedom equal to $d_F = \frac{1}{2}mp(p+1) - pk + g - \frac{1}{2}mk(k+1) - mp$ where m equals the number of groups (three in this case). To test H_F given H_k holds, the test statistic is $\chi^2_{F \cdot K} = \chi^2_F - \chi^2_k$ with degrees of freedom equal to $d_{F \cdot k} = d_F - d_k$.

If this hypothesis is found to be tenable (non-significant chi-square), we might then proceed to a more restricted structural model and test, $H_{F\Psi}$: $F_1 = F_2 = F_3, \Psi_1 = \Psi_2 = \Psi_3$. To test this we must estimate $F, \Phi_1, \Phi_2, \Phi_3$, and Ψ. Again, twice minimum F is a chi-square, χ^2_F with degrees of freedom equal to $d_{F\Psi} = \frac{1}{2}mp(p+1) - pK + g - \frac{1}{2}mK(K+1) - p$. The test statistic for $H_{F\Psi}$ is $\chi^2_{F\Psi} - \chi^2_F$

with degrees of freedom $d_{\Psi \cdot F}=d_{F\Psi}-d_F$.

If this hypothesis is also found tenable, we will proceed to test the hypothesis of equal factor patterns, factor variances-covariances, and unique score variances, $H_{F\Phi\Psi}$: $F_1 = F_2 = F_3$; $\Phi_1= \Phi_2= \Phi_3$; $\Psi_1= \Psi_2= \Psi_3$. This is a stronger hypothesis than H_Σ since H_Σ includes cases where Σ is not of the form (1-2). This hypothesis $H_{F\Phi\Psi}$ is tested directly on the basis of pooled-S, where pooled-S $= \frac{1}{n} \sum_{g=1}^{3} n_g S_g$ and $n = \sum_{g=1}^{3} n_g$. The test statistic $\chi^2_{F\Phi\Psi \cdot \Sigma} = \chi^2_{F\Phi\Psi} - \chi^2_\Sigma$ with degrees of freedom, $d_{F\Phi\Psi \cdot \Sigma} = d_{F\Phi\Psi} - d_\Sigma = \frac{1}{2}p(p+1)-pk+g-\frac{1}{2}k(k+1)-p$.

It is probably more likely to expect that H_F (identical factor patterns), H_F (equality of factor patterns, unique score variance, and the factor score variance-covariance matrices) will not be obtained. If, however, the factor patterns and factor variances proved to be statistically dissimilar over groups, then this would suggest that course content area or discipline does indeed affect teacher ratings on multidimensional instruments. In effect, comparisons of teaching abilities for faculty members within different disciplines should not be made on the basis of the factor scores alone. Perhaps, some method could be developed to control such dissimilarities in ratings which would permit limited comparisons of teachers across disciplines.

The object of the analysis in this section was to clearly demonstrate the stability of observed factor structure (representing the dimensions of teacher "skill", teacher "rapport", course "content", and course "organization/structure") over each economics class section as well as other disciplines (statistics and German). However, the actual maximum likelihood analyses necessary to test the complete factor structures could not be performed because of storage limits in the computer program. The capacity of the Jorsekog (1971) program as originally written was too small to handle the number of variables and populations in this study.

The post-course questionnaire used to obtain the current data had twenty-four specific items. Even though a total of twenty-four variables did not exceed the operational limits on Joreskog's program, the number of variables combined with the number of factors (four) and populations (four) contained in the study did. Therefore, the small internal program limits combined with an insufficient amount of time to properly expand the SIFASPF Program to handle the amount of data contained in this study were responsible for the incomplete results reported here. (The actual SIFASPF Program is very large, having over 2,000 logical steps and some

thirty-seven sub-routines which require nearly 150k CPU computer capacity to load. Consequently, rewriting the program without documentation proved essentially impossible.) The current results, then, are based upon only a subset of the original data descriptive of three teacher/course dimensions: teacher "skill", teacher-student "rapport", and course "content" coverage. There were 17 (p=17) specific items which made up these three factors. Unfortunately, several limits internal (also not discussed in published literature) to the SIFASPF Program were not compatable with one of the initial analyses (H_k) and lead to a failure of a follow-up analysis. These program and empirical failures will be described at length in the paragraphs which follow.

Maximum likelihood hypotheses of the type described earlier must be tested sequentially. For example, the test for equality of the variance-covariance matrices (H_Σ) over economics sections required some assumptions about the number of factors (k) which far exceeded the program limits. The number of factors (k) has to be set equal to the number of variables (p=17) to perform this test. The program limit on k turned out to be ten in this instance. In fact, a further reduced subset of data (and two factors) was also too large for the program with its current parameter limits. Therefore, no overall test of the equality

of the total variance-covariance matrices could be performed. This test, while extremely useful, was not critical in proceeding with other analyses to test for the optimal number of factors (k) and the equality of the factor patterns (Λg) for economics classes

Tests for the optimal number of factors (H_k) were expected to indicate that three factors were appropriate to explain the intercorrelations in the data. The results of the maximum likelihood test (H_k) proved to be highly significant ($\chi^2_k = 1357.52$, df = 396, $p<.0001$) for the four economics classes together as well as for each individual class section. This result indicated that the optimal number of factors needed to explain the specific rating item intercorrelations must be greater than three. This suggestion of more than three factorswas quite surprising since more traditional factor analytic methods (orthogonal, varimax rotated factors) had shown these three factors ("skill", "rapport", and "content") to be highly interpretable and reasonably consistent over economics class sections. However, Nunnally (1967) reports that maximum likelihood methods tend to, as a rule, over estimate the number of factors present in data obtained from any given populations. This could be a result of working with large numbers of subjects. In this case, factor loadings which are

normally small (less than .20) would become significant and, therefore, lead to an inflated number of factors necessary to explain a set of data. The reader should note that an optimal number of factors must be determined before additional hypothesis testing can be done using Joreskog's (1971) maximum likelihood procedures. However, as mentioned earlier, the program limits prohibited further analysis. Specifically, a test of the equality of the factor structures ($H_{\Lambda \cdot k}$) could not be obtained. This analysis is dependent upon the results of H_k in the following manner: the chi-square for equal factor pattern matrices ($H_{\Lambda \cdot k}$) equals chi-square H_Λ minus chi-square H_k. In addition, the lack of a set of consistent factors over economics classes would have implications concerning the expectations for consistent factors over disciplines. If variables inherent to a rater population or individual teachers affect the consistency of factors, then the affect of discipline upon the factor structure could not be assessed. This latter statement is a problem for the current analysis in addition to those already described for maximum likelihood techniques. The combination of these problems prohibited the completion of all the analyses required in this section.

In lieu of these difficulties, there were two possible

courses of action: 1) to further reduce the number of variables and factors, and proceed with additional maximum likelihood analyses; or 2) to try some alternate method to determine the similarity of factors over disciplines.

The first alternative failed, since reducing the number of factors to two still required a sufficiently large number of variables to prohibit some of the maximum likelihood analyses. Furthermore, this would mean analyzing a set of four factors, two at a time over four populations. This would produce a very large number of comparisons, which at best could only approximate in accuracy the total maximum likelihood solution. Therefore, further maximum likelihood analyses were abandoned until Joreskog's program limits have been sufficiently increased. Option two comes from Nunnally (1967, p. 356), who suggests that, rather than try to compare factor loadings, one might try cross-validating factors. That is, one might correlate factors obtained from different subjects rating the same teacher. While this method might work for the economics class section, one could not hope to determine the consistency of the factors over disciplines. The "disciplines" effect was the major focus of the analysis of this section, and so, this procedure was also abandoned.

Hopefully, the SIFASPF Program will be expanded at some time in the immediate future and the important analyses described in this section will be completed. However, the expectation of identical factor structures over identical economics sections, to say nothing of different disciplines, seems a little less likely based upon some newly discovered literature. The literature contains evidence for this statement in addition to that of the problems with over-estimation common to maximum likelihood methods. Nunnally, to be more specific, has further pointed out that factor structures may vary with rater population characteristics, such as sex, age, and education level. If this is true, then factor structures will probably vary over identical courses as well as over disciplines, due to numerous rater population characteristics. A definitve answer will have to wait until further analyses can be performed.

In conclusion, it appears that current analytic methods may not be adequate to sucessfully answer the question of whether or not course discipline will affect factor structures derived from specific rating items. However, additional evidence has been reported which suggests that such structures may not be consistent over identical courses taught by different teachers. This does not mean that the research project is without merit

or that it should not be given further consideration, but that the appropriate analyses are beyond the scope of this dissertation.

REFERENCES

Aleamoni, L.M. - A review of recent reliability and validity studies on the Illinois course evaluation questionnaire (CEQ), Research memorandum #127. Measurement and Research Division, University of Illinois, 1972 (a).

Aleamoni, L.M. - Illinois course evaluation questionnaire (CEQ) results interpretation manual. Research report #331. Measurement and Research Division, University of Illinois, 1972 (b), 1-65.

Aleamoni, L.M. - Evaluation by students to identify general instructional problems. Symposium on "A plan for the comprehensive evaluation of college teaching" at AERA. New Orleans, February 27, 1973.

Aleamoni, L.M. - The evaluation of instruction and the use of the CEQ. Research memorandum #152. Measurement and Research Division, University of Illinois, 1974 (a).

Aleamoni, L.M. - Typical faculty concerns about student evaluation of instruction, Symposium on "Methods of improving university teaching" at Technion-- Israel Institute of Technology. Haifa, Israel, March 24, 1974.

Aleamoni, L.M. - Proposed system for rewarding and improving instructional effectiveness. Technion--Israel Institute of Technology. Haifa, Israel, June, 1974 (c), 1-12.

Aleamoni, L.M. - The usefulness of student evaluations in improving college teaching, Research report #356. Measurement and Research Division, University of Illinois, 1974 (d).

Aleamoni, L.M. and Graham, M.H. - The relationship between CEQ ratings and instructors' rank, class size, and course level. Journal of Education Measurement, 1974, 11(2), 189-202.

Aleamoni, L.M. and Hexner, P.Z. - The effects of different sets of instructions on student course and instructor evaluation, Research report #339. Measurement and Research Division, University of Illinois, 1973, 1-14.

Aleamoni, L.M. and Spencer, R.E. - The Illinois course evaluation questionnaire: a description of its development and a report of some of its results. Educational and Psychological Measurement, 1973, 33, 669-684.

Aleamoni, L.M. and Yimir, M. - An investigation of the relationship between colleague ratings, student ratings, research productivity, and academic rank in rating instructional effectiveness. Journal of Educational Psychology, 1973, 64(3), 274-277.

Aleamoni, L.M. and Yimir, M. - Graduating senior ratings relationship to colleague ratings, student ratings, research productivity and academic rank in rating instructional effectivness. Research report #352. Measurement and Research Division, University of Illinois, 1974, 1-9.

Aleamoni, L.M., Yimir, M. and Maltan, J.M. - Teacher folklore and sensitivity of a course evaluation questionnaire. Psychological Reports, 1972, 31, 607-614.

Baird, L.L. - Teaching styles: an exploratory study of dimensions and effects. Journal of Educational Psychology, 1973, 64(1), 15-21.

Bausell, R.B. and Magoon, J. - Instructional methods and college students' rating of courses and instructors. The Journal of Experimental Education. 1972, 40(4), 29-33.

Bejer, I.I. and Doyle, K.O. - Generalizability of factor structures underlying student ratings of instruction. Measurement Services Center, University of Minnesota, 1974.

Bettencourt, H.O. - Student evaluations and the fickle finger of faculty fate. Unpublished review paper, University of Masschusetts, 1974.

Bock, R.D. and Bargmann, R.E. - Analysis of covariance structure. Psychometrika, 1966. 31, 507-534.

Braunstein, D.N. and Klein, G.A. - Feedback expectancy and shifts in student ratings of college faculty. Journal of Applied Psychology, 1973, 58(2), 254-258.

Carrier, M.A., Howard, G.S. and Miller, W.G. - Course evaluation: When. Journal of Educational Psychology 1974, 66(4), 609-613.

Centra, J.A. - Effectiveness of student feedback in modifying college instruction. Journal of Educational Phychology, 1973 (a), 10(1), 287-295.

Centra, J.A. - Self-ratings of college teachers: a comparison with student ratings. Journal of Educational Measurement, 1973, (b) 10(1), 287-295.

Centra, J.A. - The student as Godfather? The impact of student ratings on academia. Education Researcher, 1973, (c) 2 (10), 4-8.

Chambers, J.A. - College teachers: their effect on creativity of students. Journal of Educational Psychology, 1973, 65(3), 326-334.

Cohen, J. and Cohen, P.- Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, New Jersey: Lawrence Erlbaum Assoc., 1975.

Cohen, S.H. and Berger, W.G. - Dimensions of students' ratings of college instructors underlying subsequent achievement on course examinations. Proceedings of 78th Annual Convention of APA, 1970.

Costin, F., Greenough, W.J. and Menges, R.J. - Student ratings of college teaching: reliability, validity, and usefulness. Review of Educational Research, 1971, 41(5), 511-535.

Costin, F. and Grush, J.E. - Personality correlates of teacher-student behavior in the college classroom. Journal of Educational Psychology, 1973, 65 (1), 35-44.

Crawford, P.L. and Bradshaw, H.L. - Perception of characteristics of effective university teachers: a scaling analysis. Education and Psychological Research, 1968, 28,1079-85.

Darlington, R.B. - Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69 (3), 161-182.

Dawes, R.M. and Corrigan, B. - Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Deming, E.W. - Memorandum on Teaching. The American Statistician, February, 1972, 26(1), 47.

deWolf, V.A. - Student ratings of instruction in post secondary institutions: a comprehensive annotated bibliographyof research reported since 1968 (Vol. I), Project #264. Educational Assessment Center, University of Washington, Seattle, 1974.

Dowaliby, F.J. and Schumer, H. - Teacher-centered versus student-centered mode of college classroom instruction as related to manifest anxiety. Journal of Educational Psychology, 1973, 64(2), 125-132.

Doyle, K.O. - Student Evaluation of Instruction. Lexington, Massachusetts: D.C. Health and Co., 1975.

Doyle, K.O. - Reply to Rodin and Rodin, Research report #732F. Measurement Service Center, University of Minnesota, 1972, 1-12.

Doyle, K.O. and Whitely, S.E. - Student ratings as criteria for effective teaching. American Educational Research Journal, 1975, 11(3), 259-274.

Elmore, P.B. and LaPointe, K.A. - Effects of teacher sex and student sex on the evaluation of college instruction. Journal of Education Psychology, 1974, 66(3), 386-389.

Elmore, P.B.and LaPointe, K.A.- Effect of teacher sex, student sex and teacher warmth on the evaluation of college instructors. Journal of Educational Psychology, 1975, 67(3), 368-374.

Elton, C.F. and Rose, H.A. - Prediction of productivity from personality test scores. Journal of Educational Psychology, 1974, 66(3), 424-431.

Evans, G.T. - Transformation of factor matrices to achieve congruence. British Journal of Mathematical Psychology, 1971, 24, 22-48.

Everly, J.C. and Aleamoni, L.M. - The rise and fall of the advisor... students' attempts to evaluate their instructors. University of Illinois.

Feather, N.T. - Course evaluation and examination performance in psychology. Australian Psychologist, 1971, 6(2), 118-129.

Feather, N.T. - Dimensions of teaching effectiveness and course evaluation based upon judgements of psychology students. Australian Psychologist, 1972, 7(3), 180-190.

French-Lazovik, G.- Predictability of students' evaluations of college teachers from component ratings. Journal of Educational Psychology, 1974, 66(3), 373-385.

Frey, P.W. - Comparative judgement scaling of student course ratings. American Educational Research Journal, 1973, 10(2), 149-154.

Frey, P.W. - Student ratings of teaching: validity of several rating factors. Science, 1973, 182, 83-85.

Finkbeiner, C.T., Lathrop, J.S., and Schuerger, J.M. - Course and instructor evaluation: some dimensions of a questionnaire. Journal of Educational Psychology. 1973, 64 (2), 159-163.

Fruchter, B. - Manipulation and hypothesis-testing factor-analytic experimental designs. In G.R. Cattell (Ed.), Handbook of Multivariate Experimental Psychology. Rand McNally, 1966.

Gadzella, B.M. - College students' views and ratings of an ideal professor. College and University, 1968, 89-96.

Gessner, P.K. - Evaluation of instruction (one method to validate). Science. 1973, 180, 566-570.

Gillmore, G.M. and Brandenberg, D.C. - Would the proportion of students taking a class as a requirement affect student rating of the course? Report #347. Measurement and Research Division, University of Illinois, 1974, 1-25.

Granzin, K.L. and Painter, J.J. - A new explanation for students' course evaluation tendencies. American Educational Research Journal, 1973, 10(2), 115-124.

Grasha, A.F. - Variations in frames of reference and desirability of instructor behavior as factors influencing student ratings. Institute for Research and Training in Higher Education, University of Cincinnati, 1974, 1-30.

Grasha, A.F. - Principles and models for assessing faculty performance. Teaching/Learning Monograph Series, Volume I(2). Cincinnati, Ohio: University of Cincinnati, 1972, 1-25

Greenwood, G.E., Bridges, C.M., Ware, W.B. and McLean, J.E. Student evaluation of college teaching behaviors instrument: a factor analysis. Journal of Higher Education, 1973, 44(8), 596-604.

Guertin, W.H. and Bailey, J.P. - Introduction to Modern Factor Analysis. Ann Arbor, Michigan: Edwards Brothers, Inc., 1970.

Guttman, I. and Wilks, S.S. - Introductory Engineering Statistics. New York: John Wiley & Sons, Inc. 1965.

Harari, O. and Zedeck, S. - Development of behaviorally anchored scales for the evaluation of faculty teaching. Journal of Applied Psychology, 1973, 58(2), 261-265.

Harry, J. and Goldner, W.S. - The null relationship between teaching and research. Sociological Education. 1972. 45, 47-60.

Hartley, E.L. and Hogan, T.P. - Some additional factors in student evaluation of courses. American Educational Research Journal, 1972, 9(2), 241-250.

Hayes, J.R. - Research, teaching, and faculty fate. Science, 1971, 172, 227-230.

Hilderbrand, M. - How to recommend promotion for a mediocre teacher without actually lying. Experiment and Innovation: New Directions in Education. University of Claifornia, 1971, 10(1), 1-22.

Hogan, TlP. - Similarity of student ratings across instructors, courses, and time. Research in Higher Education, 1973, 1, 149-154.

Holmes, D.S. - The teaching assessment blank: a form for the student assessment of college instructors. The Journal of Experimental Education, 1971, 39(3), 34-38.

Jaeger, R.M. and Freijo, T.D. - Some psychometric questions in the evaluation of professors. Journal of Educational Psychology, 1974, 66(3), 416-423.

Jaeger, R.M. and Freigo, T.D. - Race and sex as concommitants of composite halo in teacher's evaluative ratings by (of) pupils. Journal of Educational Psychology, 1975, 67(2), 226-237.

Joreskog, K. G. - Factorial invariance of ability measures in groups differing in intelligence and socio-economic status. British Journal of Mathematical Psychology, 1971 (a), 24, 154-168.

Joreskog, K.G. - Simultaneous factor analysis in several populations. Psychometrika, 1971 (b), 36(4), 409-426.

Joreskog, K.G. - Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz (Ed.), Contemporary Developments in Mathematical Psychology, Vol.II, W.H. Freeman and Company, 1974.

Joreskog, K.G. - Testing a simple structure hypothesis in factor analysis. Psychometrika, 1967, 32(4), 165-177.

Joreskog, K.G. - Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32(4), 443-482.

Joreskog, K.G. - A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Kaplan, R.M. - Reflections on the Dr. Fox paradigm. Journal of Medical Education, 1974, 49-310-312.

Kennedy, W.W. - Grades expected and grades received: their relationship to students' evaluations of faculty performance. Journal of Educational Psychology, 1975, 67(1), 109-115.

Kerlinger, F.N. and Pedhazur, E.J. - Multiple Regression in Behavioral Research. New York: Holt, Rinehart, and Winston.

Kohlan, R.G. - A comparison of faculty evaluations early and later in the course. Journal of Higher Education, 1973, 587-595.

Kulik, J.A. - Student ratings of instruction. Presented at University of Massachusetts, April 3, 1974.

Kulik, J.A. and Kulik, C.C. - Student ratings of instruction. Teaching of Psychology, 1974, 1(2), 51-67.

Kulik, J.A. and McKeachie, W.J. - The evaluation of teachers in higher education. In F.N. Kerlinger (Ed.), Review of Research in Education, Vol. II, Peacock Press, 1973.

Linn, R.L., Centra, J.A., and Tucker, L. - Between, within and total group factor analysis of student ratings of instruction. Multivariate Behavioral Research, 1975, 277-288.

McGuigan, F.J. - Amount learned: an empirical basis for grading teachers and students. Teaching of Psychology, 1974, 1(1), 10-15.

McGuigan, F.J. - The G statistic: an index of amount learned. National Society for Programmed Instruction Journal, 1967, 6, 14-16.

McKeachie, W.J. - Student ratings of faculty. AAUP Bulletin, 1969, 1, 439-444.

McKeachie, W.J. - Yi-Guang Lin, and Mann, W. - Student ratings of teacher effectiveness: validity studies. American Educational Research Journal, 1971, 8(3) 435-445.

Minium, E.W. - Statistical Reasoning in Psychology and Education. New York: John Wiley and Sons, Inc., 1970.

Mulaik, S.A. - The Foundations of Factor Analysis. New York: McGraw-Hill, Inc., 1972.

Murray, H.G. - Predicting student ratings of college teaching from peer ratings of personality types. Teaching of Psychology, 1975, 2(2), 66-69.

Musella, D. and Rusch, R. - Student opinion on college teaching. Improving College and University Teaching.

Naftulin, D.H., Ware, J.E. and Donnelly, F.A. - The Dr. Fox lecture: a paradigm of educational seduction. Journal of Medical Education, 1973, 48,630-635.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, J.G., and Bent, D.H. - Statistical Package for the Social Sciences (2nd ed.). New York: McGraw-Hill, Inc. 1975.

Nord, W.R., Connelly, F. and Daignault, G. - Locus of control and aptitude test scores as predictors of academic achievement. Journal of Educational Psychology, 1974, 66(6), 956-961.

Null, E.J. and Walter, J.E. - Values of students and their ratings of a university professor. College Student Journal, 1972, 6(4), 46-51.

Nunnally, J.C. - Psychometric Theory. New York: McGraw-Hill, Inc., 1967,

Overall, J.E. and Klett, C.J. - Applied Multivariate Analysis. New York: McGraw-Hill, Inc., 1972.

Pamboukian, H.S. - Initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66(1), 52-56.

Perry, R.P. and Baumann, R.R. - Criteria for the evaluation of College teaching: Their reliability and validity. University of Toledo, 1973.

Perry, R.P, Niemi, R.R., and Jones K. - Effect of prior teaching evaluations and lecture presentations on ratings of teaching performance. Journal of Educational Psychology, 1974, 66(6), 851-856.

Pohlmann, J.T. - A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 1975, 10(1), 81-91.

Rayder, N.F. - College student ratings of instructors. The Journal of Experimental Education, 1968, 37(2), 76-81.

Rodewald, H. - Reply to Walsh's comments on Rodewald and Carroll. Perceptual and Motor Skills, 1974, 39, 1254.

Rodewald, H. and Carroll, J.L. - Problem in use of student ratings of faculty. Perceptual and Motor Skills, 1974, 38, 574.

Rodin, M., Frey, P.W., and Gessner, P.K. - Student evaluation. Science, 1975, 187, 555-559.

Rodin, M. and Rodin, B. - Student evaluation of teachers. Science, 1972, 177, 1164-1166.

Rosenshine, B. - Teacher behavior and student attitudes revisited. Journal of Educational Psychology, 1973, 65(2), 177-180.

Savicki, V., Schumer, H., and Stanfield, R.E. - Student role orientations and college dropouts. Journal of Counseling Psychology, 1970, 17(6), 559-566.

Sherman, B.R. and Blackburn, R.T. - Personal characteristics and teaching effectiveness of college faculty. Journal of Educational Psychology, 1975, 67(1), 124-131.

Stewart, C.T. and Malpass, L.F. - Estimates of achievement and ratings of instructors. The Journal of Educational Research, 1966, 59(8), 347-350.

Sullivan, A.M. and Skanes, G.R. - Validity of student evaluation of teaching and the characteristics of successful instruction. Journal of Educational Psychology, 1974, 66(4), 584-590.

Tatsuoka, M.M. - Multivariate analysis in education. In F. N. Kerlinger (Ed.), Review of Research in Education, Vol. I. Peacock Press, 1973.

Tolor, A.- Evaluation of perceived teacher effectiveness. Journal of Educational Psychology, 1973, 64(1), 98-104.

vanThillo, M. and Joreskog, K.G. - A general computer program for simultaneous factor analysis in several populations Research bulletin RM 71-5. Princeton, New Jersey: Educational Testing Service, 1970.

Wainer, H. - Estimating coefficients in linear models: it don't make no never mind. Psychological Bulletin, 1976, 83(1), 213-217.

Ware, J.E. - Return to Dr. Fox. The Chronicle of Higher Education, 1974 9(9).

Ware, J.E. and Williams, R.G. - The Dr. Fox effect: a study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 1975 (in press).

Whitefield, R.P. and Bramer, L.M. - The ills of college teaching: diagnosis and prescription. Journal of Higher Education, 1973, 53(1), 1-14.

Williams, R.G. and Ware, J.E. - Validity of student ratings of instruction under different incentive conditions: a further study of the Dr. Fox effect. Journal of Educational Psychology, 1976, 68(1), 48-56.

Zelby, L.W. - Student-faculty evaluation. Science, 1974, 183, 1267-1270

Appendix A

Subject: Preliminary Questionnaire about Student's Opinions of College Courses and Instructions

Prepared by: The Center for Instructional Resources and Improvement (CIRI): Evaluation and Research Program

Instructions:

Please provide the following information in the appropriate space on the "Op-Scan" sheet: STUDENT NUMBER & SEX (1 = female, 2 = male). Your name is not required. It is extremely important that you answer each item on the questionnaire in a thoughtful and candid manner. The data you provide will help determine the role of student opinion in decisions about faculty promotion and tenure, throughout this campus. How useful student data becomes in the University's personnel decision process depends entirely upon how seriously you respond to the questionnaire items.

The information you provide is completely confidential and will not be available to a faculty member until August 1975. At this time, faculty members will have access to the summary data only (e.g. mean response scores and standard deviations) in the form of a final formal written report. The same report will also be available to student participants.

If you have any questions pertaining to this questionnaire, please contact the Center for Instructional Resources and Improvement, Room 125A, Graduate Research Center and ask for Hal. Thank you.

Hal Bettencourt
(5-0828 or 5-0868)

1. The content area of this course is
   1. economics
   2. German
   3. math & stat

2. Your overall grade point average is:
   1. 3.40-4.00
   2. 2.80-3.39
   3. 2.20-2.79
   4. 1.60-2.19
   5. below 1.60

3. What was your grade point average for your last semester only:
   1. 3.40-4.00
   2. 2.80-3.39
   3. 2.20-2.79
   4. 1.60-2.19
   5. below 1.60

4. Mark your college:
   1. business
   2. arts and sciences
   3. education
   4. engineering
   5. other

5. Are you a:
   1. Freshman
   2. Sophomore
   3. Junior
   4. Senior
   5. Graduate

6. Is this course within your major, minor, other:
   1. major
   2. minor
   3. other

7. This course is (was)
   1. required
   2. elected

8. What grade would you feel satisfied with at semesters end?
   1. a
   2. b
   3. c
   4. d
   5. pass/fail

9. How much do you expect to learn from taking this course?
   1. much more than other similar courses
   2. slightly more than other similar courses
   3. about the same as other similar courses
   4. slightly less than other similar courses
   5. much less than other similar courses

10. How relevant is the content of this course to your expected occupation or profession:
    1. extremely relevant
    2. highly relevant
    3. somewhat relevant
    4. only remote relevance
    5. completely irrelevant

11. What type of classroom environment do you prefer to learn in:
    1. highly structured (formal lecture)
    2. mostly structured with some discussion
    3. mostly discussion with some structured lecture
    4. student centered discussion

12. Given your personal objectives for this course, what type of workload (outside of class) seems most reasonable:
    1. heavy
    2. moderately heavy
    3. average
    4. moderately low
    5. light

13. Your interest in this course is best described as:
    1. very high
    2. moderately high
    3. average
    4. moderately low
    5. very low

14. How much time do you plan to spend on this course compared to other courses you have elected (even though this course may be required and not elected):
    1. considerably more
    2. somewhat more
    3. about the same
    4. somewhat less
    5. much less

15. What class size do you learn most efficiently in:
    1. 5-15 students
    2. 16-30 students
    3. 31-45 students
    4. larger than 45
    5. class size is irrelevant

16-20.
Which fields of inquiry to you prefer?
Indicate your preference by Rank Ordering questionnaire items 16-20, where "1 = most preferred, 5 = least preferred".

16. Humanities (e.g., Language, English, Philosophy, Music, etc.)
    1. most preferred
    2. moderately high preference
    3. average preference (some desirable & undesirable courses)
    4. moderately low preference
    5. least preferred

17. Behavioral Sciences (e.g., Psychology, Sociology)
    1. most preferred
    2. moderately high preference
    3. average preference (some desirable & undesirable courses)
    4. moderately low preference
    5. least preferred

18. Social Sciences (e.g., History, Political Science, Economics)
    1. most preferred
    2. moderately high preference
    3. average preference (some desirable & undesirable courses)
    4. moderately low preference
    5. least preferred

19. Biological and Physical Sciences (e.g., Botany, Zoology)
    1. most preferred
    2. moderately high preference
    3. average preference (some desirable & undesirable courses)
    4. moderately low preference
    5. least preferred

20. Mathematics and Statistics
    1. most preferred
    2. moderately high preference
    3. average preference (some desirable & undesirable courses)
    4. moderately low preference
    5. least preferred

21. How many years of formal instruction and/or informal exposure have you had to this particular subject:
    1. no previous experience
    2. one year (i.e., two semesters) or less
    3. one to two years
    4. two to three years
    5. more than three years

22. What is your <u>average grade</u> for previous course(s) in this particular subject only:
    1. A
    2. B
    3. C
    4. D or less
    5. Not applicable

23. Overall, how effective was the instruction in your previous courses in this department (i.e., how much did the instruction facilitate learning?):
    1. generally, very facilitating
    2. moderately facilitating
    3. average facilitation
    4. below average facilitation
    5. this question does not apply to me

24. In general, students tend to fill out teacher/course evaluations in a conscientious manner:
    1. strongly agree
    2. agree
    3. no opinion
    4. disagree
    5. strongly disagree

25. Generally speaking, teacher/course evaluations are worthwhile because faculty members use this data as <u>feedback to improve</u> their teaching and courses:
    1. strongly agree
    2. agree
    3. no opinion
    4. disagree
    5. strongly disagree

Subject: Final Questionnaire about Student's Opinions of College Courses and Instruction

Prepared by: The Center for Instructional Resources and Improvement (CIRI): Evaluation and Research Program

Instructions:

Please provide the following information in the appropriate space on the "Op-Scan" sheet: STUDENT NUMBER & SEX (1 = female, 2 = male). Your name is not required. It is extremely important that you answer each item on the questionnaire in a thoughtful and candid manner. The data you provide will help determine the role of student opinion in decisions about faculty promotion and tenure throughout this campus. How useful student data becomes in the University's personnel decision process depends entirely upon how seriously you respond to the questionnaire items.

The information you provide is completely confidential and will not be available to a faculty member until August 1975. At this time, faculty members will have access to the summary data only (e.g., mean response scores and standard deviations) in the form of a final formal written report. The same report will also be available to student participants.

If you have any questions pertaining to this questionnaire, please contact the Center for Instructional Resources and Improvement, Room 125A, Graduate Research Center and ask for Hal. Thank you.

Hal Bettencourt
(5-0828 or 5-0868)

1. The content area of this course is:

    1. economics
    2. German
    3. Math and statistics

2. What grade do you expect to get in this course?

    1. A
    2. B
    3. C
    4. D or less
    5. Pass/Fail

3. Did you learn as much as you anticipated in this course?

    1. much more than expected
    2. more than expected
    3. about the same as expected
    4. less than expected
    5. much less than expected

4. How relevant is the content of the course to your expected occupation or profession?

    1. extremely relevant
    2. highly relevant
    3. somewhat relevant
    4. only remote relevance
    5. completely irrelevant

5. Given your personal objectives, the work load in this course was:

    1. heavy
    2. moderately heavy
    3. average
    4. moderately low
    5. light

6. Your interest in this course is best described as:

    1. very high
    2. moderately high
    3. average
    4. moderately low
    5. very low

7. How much time did you spend on this course compared to your elected courses?

    1. considerably more
    2. somewhat more
    3. about the same
    4. somewhat less
    5. much less

8. In general, you would rate the teacher's enthusiasm about lecture presentations as:

   1. very high
   2. moderately high
   3. average
   4. moderately low
   5. very low

9. In general, how entertaining did your teacher make lecture presentations?

   1. high entertainment value
   2. moderate entertainment value
   3. average entertainment value
   4. below average entertainment value
   5. no entertainment value

10. Overall, the instructor's general teaching ability was:

    1. excellent
    2. very good
    3. average
    4. below average
    5. poor

11. Overall, the instructor's attitude toward (rapport with) students was:

    1. excellent
    2. very good
    3. average
    4. below average
    5. poor

12. Overall, how effective was the course at meeting specified objectives (i.e., content area taught)?

    1. excellent coverage
    2. very good coverage
    3. average coverage
    4. below average coverage
    5. poor coverage

13. Overall, how would you rate the organization of this course?

    1. excellent
    2. very good
    3. average
    4. below average
    5. poor

Specific course rating items: Questions 14 thru 40 below should be rated using the following scale.

1. strongly agree
2. agree
3. uncertain
4. disagree
5. strongly disagree

14. Class presentations were well organized.

15. The instructor was well prepared for lectures.

16. The instructor used student questions as a source for discovering points of confusion.

17. The instructor imparted essential material that was not contained in the text.

18. I feel I profited from the out-of-class assignments.

19. This course was worthwhile because the content is directly applicable to your future occupation.

20. You were never hesitant to ask questions in this course.

21. Exams stress conceptual understanding.

22. The instructor seemed to know whether the class was following his/her presentations in class.

23. You would take another course with similar content even if you didn't have to.

24. The instructor encouraged students to ask questions during class.

25. The course material appeared to be presented in logical content units.

26. In general, this course was taught well.

27. The instructor frequently told students when they were doing particularly well.

28. The instructor's presentations made the subject matter interesting.

29. The instructor was concerned about each student's progress.

30. The instructor organized this course so that the course material related to a wider body of knowledge.

31. Homework assignments were interesting and stimulating.

32. The instructor was available for consultation with students.

33. The instructor showed a genuine interest in teaching this class.

34. This course was valuable because it increased my general knowledge.

35. The instructor clearly specified course objectives.

36. The grading procedures fairly indicated each student's performance.

37. Because of this course you have developed an increased appreciation for the subject area.

38. The instructor expressed ideas clearly.

39. Exams adequately covered the text material.

40. The instructor's voice was clear and understandable.

To what extent do the following factors enter into your overall ratings of a course? Please use the following scale to rate each item 41 thru 45.

1. extremely influential
2. highly influential
3. moderately influential
4. little influence
5. no influence at all

41. Fairness of exams

42. Teacher's rapport with students

43. Entertaining lecture presentations

44. Amount you have learned by taking the course

45. The teacher's ability to motivate your interest in the course

Questions about your attitudes and goals:

On every college or university campus students hold a variety of attitudes about their own purposes and goals while at college. Such an attitude might be thought of as a personal philosophy of higher education. Below are descriptive statements of five such "personal philosophies" which there is reason to believe are quite prevalent on American college campuses. As you read each statement, think about how well it describes your own philosophy. Then on the answer sheet mark the number that indicates how much each philosophy is like or unlike you (i.e., Philosophy A is number 46 on the answer sheet. Underneath it, darken a number 1-5 depending

upon how much Philosophy A is like your own.).

46. Philosophy A. Education as preparation for an occupation. Students are committed to particular fields of study and are in college to get training for careers in their chosen fields.

like me 1 2 3 4 5 unlike me

47. Philosophy B. Interest in ideas, pursuit of knowledge, and development of the intellect are emphasized. Students are seriously involved in course work or independent study beyond the minimum requirements.

like me 1 2 3 4 5 unlike me

48. Philosophy C. Extra-curricular activities, living-group functions, athletics, social life, regarding friendships, and school spirit are emphasized.

like me 1 2 3 4 5 unlike me

49. Philosophy D. An individual way of life is the goal. Awareness is sought through freedom to explore all emotional and physical sensations. This awareness is often in conflict with the values of society.

like me 1 2 3 4 5 unlike me

50. Philosophy E. Committment to social responsibilities is emphasized. Since their values are often in conflict with society, students are willing to work hard, and to even make sacrifices in order to change and reform society at large.

like me 1 2 3 4 5 unlike me

Now, rank order the five statements above according to the accuracy with which each portrays your own point of view. So on the answer sheet darken in a different number 1-5 for each philosophy according to its accuracy (1=most accurate, 2=second most accurate, 3=third most accurate, 4=fourth most accurate, 5=least accurate).

51. Philosophy A
52. Philosophy B
53. Philosophy C
54. Philosophy D
55. Philosophy E

Appendix B

Table B-1: Means and Standard Deviations for Pre- and Post-Course Questionnaire Items; Economics classes E1 through E4

Pre-Course Items		E1	E2	E3	E4
(2)[1]	M	1.960	2.303	2.261	2.158
	S	0.864	0.891	0.891	0.829
(3)	M	1.935	2.257	2.215	2.191
	S	0.859	1.020	0.940	0.898
(4)	M	1.971	2.774	3.044	1.933
	S	1.343	1.685	1.735	1.482
(5)	M	1.981	2.116	1.978	1.732
	S	1.073	0.895	0.943	0.885
(6)	M	1.917	1.986	2.039	1.726
	S	0.964	0.969	0.974	0.972
(7)	M	1.351	1.333	1.326	1.245
	S	0.479	0.487	0.470	0.431
(8)	M	1.507	1.486	1.580	1.455
	S	0.802	0.612	0.746	0.657
(9)	M	1.902	2.324	2.539	2.498
	S	0.786	0.781	0.743	0.785
(10)	M	2.224	2.349	2.558	2.380
	S	0.917	0.884	0.909	0.796
(11)	M	2.284	2.214	2.320	2.158
	S	0.679	0.669	0.705	0.664
(12)	M	2.888	2.850	3.044	2.967
	S	0.611	0.696	0.706	0.575
(13)	M	1.985	2.313	2.707	2.364
	S	0.787	0.842	0.898	0.695
(14)	M	2.727	2.823	2.768	2.928
	S	0.710	0.709	0.838	0.700
(15)	M	2.956	2.776	2.580	2.742
	S	1.649	1.695	1.585	1.581

(continued on next page)

1---Refer to Appendix A for item meanings and response options.

Table B-1: (continued)

			E1	E2	E3	E4
Pre-Course Items	(16)	M	3.181	3.199	3.208	3.348
		S	1.115	1.118	1.118	1.073
	(17)	M	2.873	2.849	2.903	2.923
		S	1.011	0.992	0.990	0.955
	(18)	M	2.346	2.418	2.818	2.620
		S	0.976	1.029	1.020	1.000
	(19)	M	3.436	2.938	2.806	3.308
		S	1.244	1.391	1.355	1.216
	(20)	M	3.387	3.130	3.087	2.749
		S	1.343	.1250	1.376	1.335
	(21)	M	1.825	1.616	1.506	1.966
		S	0.698	0.896	0.758	0.783
	(22)	M	2.620	3.741	3.849	2.362
		S	1.574	1.652	1.593	1.379
	(23)	M	3.221	3.993	4.041	2.870
		S	1.457	1.498	1.424	1.350
	(24)	M	2.439	2.490	2.322	2.704
		S	0.892	0.944	0.831	1.061
	(25)	M	2.867	2.841	2.500	2.637
		S	1.009	1.025	0.950	1.030
Post-Course Items	(2)	M	1.916	1.611	2.169	1.882
		S	0.948	0.609	0.908	0.835
	(3)	M	2.545	2.688	3.264	2.810
		S	0.850	0.716	0.917	0.825
	(4)	M	2.401	2.679	2.873	2.592
		S	0.949	0.849	0.988	0.887
	(5)	M	3.068	2.945	2.968	2.980
		S	0.671	0.692	0.680	0.612

(continued on next page)

Table B-1: (continued)

		E1	E2	E3	E4
(6)	M	2.130	2.798	3.159	2.484
	S	0.943	1.070	1.113	1.014
(7)	M	3.156	.103	3.328	3.0202
	S	0.790	0.868	0.840	0.844
(8)	M	1.443	1.743	2.260	1.974
	S	0.714	0.763	0.910	0.752
(9)	M	1.411	2.303	3.346	2.572
	S	0.608	0.977	1.130	0.939
(10)	M	1.438	2.110	2.937	2.217
	S	0.706	0.936	1.037	0.861
(11)	M	1.813	2.617	2.516	2.179
	S	0.742	1.034	0.978	0.872
(12)	M	1.990	2.339	2.792	2.191
	S	0.793	0,819	0.796	0.787
(13)	M	2.083	2.193	2.855	2.294
	S	0.788	0.713	0.862	0.826
(14)	M	1.547	1.927	2.512	1.824
	S	0.549	0.766	0.999	0.699
(15)	M	1.359	1.606	2.055	1.667
	S	0.513	0.593	0.848	0.585
(16)	M	2.105	2.376	2.780	3.020
	S	0.934	1.034	1.101	0.966
(17)	M	1.990	2.417	2.864	2.620
	S	0.838	0.908	0.978	0.960
(18)	M	2.798	3.067	3.008	2.613
	S	1.050	1.026	1.203	1.035
(19)	M	2.277	2.807	2.882	2.474
	S	0.974	1.093	1.245	0.996

Post-Course Items

(continued on next page)

Table B-1: (continued)

		E1	E2	E3	E4
Post-Course Items (20)	M	2.735	3.731	3.449	3.007
	S	1.018	1.204	1.135	1.103
(21)	M	2.238	1.817	2.260	2.320
	S	0.990	0.830	1.100	1.062
(22)	M	2.309	2.587	2.969	2.510
		0.823	1.107	1.076	0.933
(23)	M	2.615	3.239	3.449	3.105
	S	1.129	1.246	1.226	1.202
(24)	M	2.347	2.651	3.228	3.073
	S	0.870	1.125	1.286	1.014
(25)	M	1.797	2.037	2.433	2.105
	S	0.619	0.666	0.981	0.862
(26)	M	1.568	2.092	2.882	2.124
	S	0.668	0.948	1.159	0.955
(27)	M	3.478	3.028	3.897	3.539
	S	0.826	1.134	0.847	0.955
(28)	M	1.563	2.404	3.167	2.477
	S	0.668	1.001	1.094	0.994
(29)	M	3.258	3.255	3.614	3.315
	S	0.911	0.926	0.976	0.987
(30)	M	2.138	2.385	2.748	2.474
	S	0.827	0.891	0.891	0.829
(31)	M	3.108	3.702	3.698	3.020
	S	1.050	0.974	0.957	1.089
(32)	M	2.566	2.569	2.551	2.523
	S	0.864	0.809	0.888	0.923
(33)	M	1.568	2.009	2.102	1.967
	S	0.611	0.822	0.853	0.803

(continued on next page)

Table B-1: (continued)

		E1	E2	E3	E4
(34)	M	1.581	2.110	2.480	2.066
	S	0.634	0.750	1.075	0.884
(35)	M	1.801	2.312	2.638	2.280
	S	0.762	0.813	0.965	0.804
(36)	M	3.242	2.722	3.102	2.907
	S	1.091	1.012	1.133	1.070
(37)	M	1.958	2.495	2.827	2.480
	S	0.902	1.024	1.196	1.091
(38)	M	1.663	2.321	2.984	2.133
	S	0.644	0.881	1.182	0.833
(39)	M	2.569	2.120	2.722	2.543
	S	1.075	0.806	1.071	1.124
(40)	M	1.316	1.587	1.850	1.881
	S	0.466	0.612	0.691	0.774
(41)	M	2.495	2.222	2.472	2.405
	S	1.116	0.941	1.028	0.982
(42)	M	2.348	2.398	2.744	2.718
	S	0.929	0.842	0.975	1.097
(43)	M	1.835	2.426	2.864	2.570
	S	0.827	0.919	0.962	0.953
(44)	M	1.717	2.074	2.240	2.094
	S	0.703	0.904	1.066	1.002
(45)	M	1.871	2.324	2.496	2.349
	S	0.848	1.092	1.126	1.039

Post-Course Items

Table B-2: Percent Response Rate by Item Category for Pre- and Post-Course Questionnaires; Economics classes E1 through E4

	E1	E2	E3	E4
Sample size = N	295	182	216	269
Item #2				
1.	34.2	19.0	21.1	20.7
2.	40.2	40.8	40.4	49.3
3.	21.6	31.7	29.8	24.1
4.	3.5	7.7	8.7	5.4
5.	0.5	0.7	0.0	0.5
Item #3				
1.	35.7	25.7	26.6	22.5
2.	39.2	37.9	34.2	45.1
3.	21.6	22.9	30.4	23.5
4.	3.0	12.1	8.9	8.3
5.	0.5	1.4	0.0	0.5
Item #4				
1.	50.2	30.1	28.7	62.7
2.	30.7	32.2	23.2	16.7
3.	3.4	0.0	1.1	0.0
4.	2.9	5.5	8.8	5.7
5.	12.7	32.2	38.1	14.8
Item #5				
1.	45.1	28.6	38.7	52.2
2.	24.3	36.7	30.4	25.8
3.	18.4	29.9	26.5	19.1
4.	11.7	4.1	3.3	2.4
5.	0.5	0.7	1.1	0.5
Item #6				
1.	50.7	45.3	43.6	61.1
2.	6.8	12.2	9.9	7.7
3.	42.4	41.9	45.9	29.8

Table B-2: (continued)

	E1	E2	E3	E4
Item #7				
1.	64.9	67.3	67.4	75.5
2.	35.1	32.0	32.6	24.5
Item #8				
1.	59.5	55.4	50.8	59.3
2.	36.6	41.9	44.8	38.8
3.	0.5	2.0	2.2	0.8
4.	0.5	0.0	0.0	0.0
5.	2.9	0.7	2.2	1.4
Item #9				
1.	34.6	16.6	10.6	12.4
2.	41.5	36.6	28.3	30.1
3.	23.4	45.5	58.9	53.6
4.	0.0	0.7	1.1	2.9
5.	0.5	0.7	1.1	1.0
Item #10				
1.	22.4	17.8	12.7	15.9
2.	42.0	37.7	33.1	34.1
3.	27.8	37.7	41.1	46.2
4.	6.3	5.5	11.0	3.8
5.	1.5	1.4	1.7	0.0
Item #11				
1.	8.3	11.7	8.3	13.4
2.	59.3	57.2	56.9	59.3
3.	27.3	29.0	29.3	25.4
4.	4.4	2.1	5.5	1.9
5.	0.0	0.0	0.0	0.0

Table B-2: (cOntinued)

	E1	E2	E3	E4
Item #12				
1.	0.5	1.4	1.1	0.5
2.	21.4	27.2	13.8	14.4
3.	68.9	57.8	70.2	75.6
4.	7.3	12.2	9.4	7.2
5.	1.9	1.4	5.5	2.4
Item #13				
1.	26.2	16.3	6.6	6.7
2.	53.4	44.2	34.3	55.5
3.	17.0	31.3	45.9	33.0
4.	2.4	8.2	8.3	4.3
5.	1.0	0.0	5.0	0.5
Item #14				
1.	2.9	2.7	8.3	2.9
2.	33.2	26.5	22.7	19.1
3.	52.7	57.1	54.7	60.8
4.	10.7	12.9	12.7	16.7
5.	0.5	0.7	1.7	0.5
Item #15				
1.	23.9	29.9	32.0	24.9
2.	29.8	32.0	32.0	36.8
3.	9.3	3.4	8.3	6.7
4.	1.0	0.0	1.1	2.4
5.	36.1	34.7	26.5	29.2
Item #16				
1.	7.8	5.5	6.2	5.3
2.	15.7	19.2	18.0	13.0
3.	42.2	43.8	42.1	40.1
4.	19.1	13.0	16.3	24.6
5.	15.2	18.5	17.4	16.9

Table B-2: (continued)

	E1	E2	E3	E4
Item #17				
1.	8.3	9.6	6.3	2.9
2.	27.8	24.0	27.8	34.6
3.	37.6	43.2	42.6	36.1
4.	21.0	18.5	15.9	20.2
5.	5.4	4.8	7.4	6.3
Item #18				
1.	22.4	21.9	9.7	14.4
2.	32.7	30.8	27.8	29.3
3.	34.1	32.9	39.2	39.4
4.	9.3	12.3	17.6	13.5
5.	1.5	2.1	5.7	3.4
Item #19				
1.	9.3	23.3	23.4	7.7
2.	13.7	13.0	19.4	20.7
3.	24.0	26.7	23.4	24.0
4.	29.9	20.5	20.6	28.4
5.	23.0	16.4	13.1	19.2
Item #20				
1.	10.8	10.3	17.3	20.8
2.	17.2	24.0	17.9	27.5
3.	23.0	25.3	23.7	22.2
4.	20.6	23.3	20.8	15.0
5.	28.4	17.1	20.2	14.5
Item #21				
1.	29.5	56.8	60.3	21.6
2.	62.5	32.2	32.8	67.8
3.	5.0	5.5	4.0	5.8
4.	2.0	3.4	1.7	1.9
5.	1.0	2.1	1.1	2.9

Table B-2: (continued)

	E1	E2	E3	E4
Item #22				
1.	31.0	16.1	13.4	31.9
2.	30.0	16.8	16.3	34.8
3.	11.5	4.9	5.8	15.0
4.	1.0	1.4	1.2	1.9
5.	26.5	60.8	63.4	16.4
Item #23				
1.	15.6	12.4	9.9	21.3
2.	22.1	9.7	9.9	20.3
3.	15.1	8.3	9.9	22.7
4.	19.1	5.5	7.0	21.7
5.	28.1	64.1	63.4	14.0
Item #24				
1.	10.1	6.9	9.9	8.7
2.	52.5	58.6	60.8	44.7
3.	21.2	16.6	17.0	18.4
4.	15.7	14.5	11.7	23.8
5.	0.5	3.4	0.6	4.4
Item #25				
1.	6.6	5.5	9.4	9.3
2.	32.7	37.9	50.6	45.1
3.	33.7	30.3	24.1	22.5
4.	21.4	19.3	12.4	18.6
5.	5.6	6.9	3.5	4.4

Table B-2: (continued)

	E1	E2	E3	E4
Item #2				
1.	35.1	45.4	22.6	32.7
2.	47.6	48.1	46.8	52.3
3.	12.6	6.5	24.2	11.8
4.	0.0	0.0	4.0	0.7
5.	4.7	0.0	2.4	2.6
Item #3				
1.	8.4	4.6	3.2	4.6
2.	41.9	32.1	12.8	29.4
3.	38.7	53.2	48.0	48.4
4.	8.9	10.1	24.4	15.7
5.	2.1	0.0	9.6	2.0
Item #4				
1.	17.2	8.3	8.7	9.9
2.	39.6	31.2	25.4	36.2
3.	30.7	45.9	39.7	40.8
4.	10.9	13.8	22.2	11.2
5.	1.6	0.9	4.0	2.0
Item #5				
1.	1.0	1.8	1.6	2.0
2.	13.0	19.3	17.5	18.8
3.	67.2	63.3	65.9	74.5
4.	15.6	13.8	12.7	9.8
5.	3.1	1.8	2.4	2.0
Item #6				
1.	27.6	11.0	4.8	16.3
2.	40.6	26.6	24.6	37.3
3.	25.0	43.1	36.5	32.0
4.	4.7	10.1	18.3	10.5
5.	2.1	9.2	15.9	3.9

Table B-2: (continued)

	E1	E2	E3	E4
Item #7				
1.	2.1	4.7	0.8	2.6
2.	12.0	15.0	13.6	21.1
3.	60.4	49.5	45.6	53.0
4.	19.3	27.1	32.0	17.9
5.	6.3	3.7	8.0	5.3
Item #8				
1.	66.1	44.0	19.7	29.4
2.	25.5	38.5	44.9	43.8
3.	6.8	16.5	26.8	26.8
4.	1.0	0.9	7.1	0.0
5.	0.5	0.0	1.6	0.0
Item #9				
1.	64.6	20.2	3.1	11.2
2.	30.2	43.1	21.3	37.5
3.	4.7	25.7	34.6	37.5
4.	0.5	8.3	19.7	10.5
5.	0.0	2.8	21.3	3.3
Item #10				
1.	65.6	27.5	7.9	20.4
2.	27.6	43.1	25.2	44.7
3.	4.7	22.0	40.2	28.3
4.	1.6	5.5	18.9	5.9
5.	0.5	1.8	7.9	0.7
Item #11				
1.	36.5	12.1	12.7	23.2
2.	47.9	37.4	41.3	42.4
3.	13.5	32.7	31.7	28.5
4.	2.1	12.1	10.3	5.3
5.	0.0	5.6	4.0	0.7

Table B-2: (continued)

	E1	E2	E3	E4
Item #12				
1.	27.6	12.8	3.2	16.4
2.	50.0	48.6	32.0	53.3
3.	18.2	31.2	49.6	26.3
4.	4.2	6.4	12.8	2.6
5.	0.0	0.9	2.4	1.3
Item #13				
1.	22.4	16.5	4.0	16.3
2.	51.0	48.6	29.8	44.4
3.	22.9	33.9	46.0	33.3
4.	3.1	0.9	16.9	5.2
5.	0.5	0.0	3.2	0.7

Table B-3: Means and Standard Deviations for Pre- and Post-Course Questionnaire Items; Grouped Data, Statistics (S), German (G), and Economics (E)

Pre-Course Items

		S	G	E
(2)[1]	M	2.264	1.720	2.155
	S	0.883	0.831	0.874
(3)	M	2.170	1.697	2.137
	S	0.932	0.864	0.930
(4)	M	1.149	2.260	2.381
	S	0.587	0.818	1.626
(5)	M	2.682	1.805	1.937
	S	0.774	1.159	0.964
(6)	M	1.743	2.679	1.907
	S	0.948	0.712	0.975
(7)	M	1.061	1.416	1.312
	S	0.267	0.522	0.466
(8)	M	1.899	1.808	1.506
	S	1.177	1.207	0.713
(9)	M	2.662	2.346	2.309
	S	0.923	0.835	0.816
(10)	M	2.878	3.346	2.374
	S	0.903	0.978	0.882

Pre-Course Items

		S	G	E
(11)	M	2.184	2.385	2.244
	S	0.652	0.688	0.681
(12)	M	2.796	2.833	2.941
	S	0.776	0.746	0.646
(13)	M	2.871	2.372	2.332
	S	0.981	0.899	0.843
(14)	M	2.558	2.692	2.813
	S	0.923	1.048	0.743
(15)	M	2.054	1.949	2.768
	S	1.198	1.502	1.626
(16)	M	3.288	2.455	3.238
	S	1.220	1.165	1.104
(17)	M	2.993	2.668	2.890
	S	1.050	1.003	0.985
(18)	M	2.660	3.184	2.551
	S	1.610	0.920	1.019
(19)	M	3.401	2.227	3.150
	S	1.145	1.311	1.317

1---Refer to Appendix A for item meanings and response options.

(continued on next page)

Table B-3: (continued)

			S	G	E
Pre-Course Items	(20)	M	2.844	3.640	3.084
		S	1.198	1.362	1.350
	(21)	M	1.567	2.107	1.747
		S	0.772	0.746	0.798
	(22)	M	3.924	1.667	3.061
		S	1.519	0.963	1.674
	(23)	M	3.542	2.067	3.470
		S	1.379	1.070	1.511
	(24)	M	2.467	2.392	2.497
		S	1.007	1.018	0.949
	(25)	M	2.882	2.521	2.709
		S	1.174	0.988	1.014
Post-Course Items	(2)	M	2.661	1.940	1.905
		S	1.411	1.140	0.871
	(3)	M	2.625	2.821	2.798
		S	1.041	0.777	0.875
	(4)	M	2.874	3.328	2.606
		S	0.935	0.927	0.938
Post-Course Items	(5)	M	2.670	2.373	3.000
		S	0.606	0.775	0.662
	(6)	M	3.072	2.597	2.572
		S	1.085	0.986	1.093
	(7)	M	2.732	2.530	3.148
		S	0.958	1.011	0.835
	(8)	M	1.955	1.896	1.818
		S	0.999	0.890	0.837
	(9)	M	2.694	2.269	2.307
		S	1.197	0.898	1.153
	(10	M	2.250	2.060	2.097
		S	1.263	0.868	1.029
	(11)	M	2.304	1.851	2.212
		S	1.169	0.783	0.944
	(12)	M	2.205	2.030	2.282
		S	1.015	0.887	0.848
	(13)	M	2.313	2.134	2.325
		S	1.040	0.886	0.850

(continued on next page)

Table B-3: (continued)

Post-Course Items

		S	G	E
(14)	M	1.938	1.864	1.902
	S	0.989	0.699	0.824
(15)	M	1.821	1.612	1.639
	S	1.006	0.602	0.680
(16)	M	2.179	1.802	2.542
	S	1.015	0.764	1.079
(17)	M	2.500	2.299	2.424
	S	1.090	0.905	0.974
(18)	M	2.345	2.431	2.845
	S	1.062	1.075	1.089
(19)	M	2.892	3.313	2.561
	S	1.194	1.144	1.093
(20)	M	2.360	1.806	3.151
	S	1.212	0.857	1.162
(21)	M	2.343	2.388	2.185
	S	1.185	1.114	1.021
(22)	M	2.577	2.030	2.559
	S	1.290	0.852	0.995

Post-Course Items

		S	G	E
(23)	M	3.658	2.955	3.043
	S	1.187	1.221	1.233
(24)	M	2.171	1.970	2.789
	S	0.862	0.738	1.117
(25)	M	2.108	2.075	2.062
	S	1.039	0.926	0.816
(26)	M	2.198	1.985	2.100
	S	1.348	0.879	1.036
(27)	M	3.352	2.433	3.504
	S	1.092	1.018	0.968
(28)	M	2.708	2.239	2.310
	S	1.280	0.971	1.098
(29)	M	2.717	2.015	3.352
	S	1.040	0.807	0.957
(30)	M	2.731	2.582	2.407
	S	1.054	0.907	0.881
(31)	M	3.181	3.552	3.324
	S	1.142	1.105	1.071

(continued on next page)

Table B-3: (continued)

Post-Course Items

		S	G	E
(32)	M	1.867	1.672	2.552
	S	0.867	0.683	0.873
(33)	M	1.788	1.746	1.873
	S	0.972	0.785	0.790
(34)	M	2.291	2.284	2.005
	S	1.054	1.027	0.898
(35)	M	2.350	2.209	2.206
	S	1.082	0.826	0.886
(36)	M	2.515	2.164	3.026
	S	1.298	0.979	1.095
(37)	M	2.621	2.299	2.387
	S	1.230	1.073	1.093
(38)	M	2.369	1.955	2.201
	S	1.237	0.806	1.001
(39)	M	1.990	2.000	2.511
	S	0.965	0.985	1.059
(40)	M	1.621	1.746	1.633
	S	0.688	0,927	0.680

Post-Course Items

		S	G	E
(41)	M	2.029	2.358	2.415
	S	0.975	1.055	1.033
(42)	M	2.305	1.866	2.541
	S	0.962	0.851	0.985
(43)	M	2.781	2.463	2.365
	S	1.168	1.005	0.990
(44)	M	2.295	1.851	1.998
	S	0.990	0.803	0.932
(45)	M	2.257	2.179	2.220
	S	1.029	1.043	1.040

Table B-4: Percent Pesponse Rate by Item Category for Pre- and Post-Course Questionnaires; Grouped Data, Statistics (S), German (G), and German (G)

	S[1]	G	E
Item #2			
1.	21.6	45.3	24.3
2.	37.2	42.7	43.0
3.	35.1	8.0	26.2
4.	5.4	2.7	6.1
5.	0.7	1.3	0.4
Item #3			
1.	25.9	51.3	27.8
2.	40.8	31.6	39.5
3.	24.5	14.5	24.4
4.	8.2	1.3	7.7
5.	0.7	1.3	0.6
Item #4			
1.	92.6	1.3	44.5
2.	2.7	87.0	25.2
3.	2.7	2.6	1.2
4.	1.4	2.6	5.7
5.	0.7	6.5	23.3
Item #5			
1.	1.4	61.0	42.3
2.	43.9	11.7	28.7
3.	42.6	15.6	22.9
4.	9.5	9.1	5.5
5.	2.7	2.6	0.7
Item #6			
1.	60.8	14.1	50.8
2.	4.1	3.8	8.9
3.	35.1	82.1	39.6
Item #7			
1.	94.6	59.7	69.0
2.	94.6	59.7	69.0

(continued on next page)

1---Sample sizes: S = 145; G = 78; E = 743.

Table B-4: (continued)

	S	G	E
Item #8			
1.	45.9	52.6	56.5
2.	37.8	34.6	40.2
3.	5.4	2.6	1.2
4.	2.0	0.0	0.1
5.	8.8	10.3	1.9
Item #9			
1.	10.1	17.9	18.9
2.	29.1	33.3	34.1
3.	50.7	46.2	49.9
4.	4.7	1.3	1.2
5.	5.4	1.3	0.8
Item #10			
1.	6.8	2.6	17.3
2.	23.6	12.8	36.8
3.	48.0	47.4	38.2
4.	18.2	21.8	6.6
5.	3.4	15.4	1.1
Item #11			
1.	8.8	6.4	10.4
2.	68.0	53.8	58.3
3.	19.7	34.6	27.7
4.	2.7	5.1	3.5
5.	0.7	0.0	0.0
Item #12			
1.	4.1	3.8	0.8
2.	25.9	24.4	18.7
3.	60.5	57.7	68.9
4.	5.4	12.8	8.7
5.	4.1	1.3	2.8
Item #13			
1.	4.1	16.7	14.0
2.	34.0	39.7	47.5
3.	41.5	34.6	31.4
4.	17.6	17.7	5.5
5.	8.8	1.3	1.6

Table B-4: (continued)

	S	G	E
Item #14			
1.	13.6	12.8	4.2
2.	32.0	33.3	25.3
3.	40.8	28.2	56.3
4.	12.2	23.1	13.3
5.	1.4	2.6	0.8
Item #15			
1.	34.7	60.3	27.4
2.	48.3	20.5	32.7
3.	4.8	1.3	7.1
4.	1.4	0.0	1.2
5.	10.9	17.9	31.5
Item #16			
1.	6.2	28.6	6.3
2.	21.9	18.2	16.2
3.	32.2	37.7	41.9
4.	16.4	10.4	18.8
5.	23.3	5.2	16.9
Item #17			
1.	8.8	10.4	6.5
2.	21.1	33.8	29.0
3.	39.5	27.7	39.5
4.	23.1	13.0	19.0
5.	7.5	5.0	6.0
Item #18			
1.	12.2	2.6	17.0
2.	32.7	17.1	30.2
3.	36.1	48.7	36.6
4.	15.0	22.4	13.1
5.	4.1	9.2	3.1
Item #19			
1.	6.8	44.0	15.0
2.	14.3	16.0	16.9
3.	29.3	18.7	24.4
4.	31.3	16.0	25.4
5.	18.4	5.3	18.3

Table B-4: (continued)

	S	G	E
Item #20			
1.	10.9	9.3	15.1
2.	34.0	10.7	21.6
3.	28.6	28.0	23.4
4.	12.9	10.7	19.6
5.	13.6	41.3	20.3
Item #21			
1.	62.3	14.7	40.1
2.	28.1	68.0	50.8
3.	6.8	9.3	5.1
4.	2.1	8.0	2.2
5.	0.7	0.0	1.8
Item #22			
1.	10.3	56.0	24.1
2.	17.2	29.3	25.5
3.	4.8	10.7	9.8
4.	4.8	0.0	1.4
5.	62.5	4.0	39.2
Item #23			
1.	9.7	41.3	15.2
2.	16.0	21.3	16.2
3.	21.5	28.0	14.7
4.	16.0	8.0	14.2
5.	36.8	1.3	39.7
Item #24			
1.	12.4	13.5	9.0
2.	49.7	55.4	53.5
3.	19.3	13.5	18.5
4.	15.2	13.5	16.8
5.	3.4	4.1	2.2
Item #25			
1.	12.5	11.0	7.8
2.	27.8	46.6	41.5
3.	28.5	26.0	27.6
4.	21.5	12.3	18.0
5.	9.7	4.1	5.0

Appendix C

TABLE 1A

Apparent Similarity of Dimensions in Factor Analyses of Student Evaluation of College Teaching*

Study	Skill	Rapport	Structure	Overload	Other
Smalzeid & Remmers (1943)	Maturity	Empathy			
Creager (1950)	Professional Impression	Rapport			
Bendig (1954)	Instructor Competence	Instructor Empathy			+1
Gibb (1955)	Communication	Friendly-Democratic	Organization	Academic Emphasis	
Isaacson et al (1964)	Skill	Rapport	Structure	Overload	+2
Solomon (1966)	Energy vs Lethargy	Lecturing vs Student Participation	Control vs Permissiveness	Criticism vs Tolerance	+6
Turner (1970)	Exciting, Humorous, Stimulating	Approachable, Warm, Cheerful	Penetrating, Clear, Focused	Prepared, Probing, Demanding	+2
Deshpande et al (1970) 2nd Order Factors	Stimulation	Affective Merit	Cognitive Merit	Stress	
Hartley & Hogan (1972)	Overall Evaluation	Student-Teacher Interaction	Structure or Organization	Load or Difficulty	+3 (42%)
Frey (1973)	Teacher's Presentation	Teacher Accessibility	Organization, Planning	Workload	+1 (67.5%)

*This table taken from Kulik & McKeachie (1973), page 4a.

TABLE 1B

Study	Skill	Rapport	Structure	Overload	General Course Attitude	Others
Holmes (1971)	Instruction Presentation (24%)	Interaction Evaluation (11.7%)		Test Clarity (10.5%)	Student Stimulation (16.8%)	+1 (63.3%)**
McKeachie, Lin & Mann (1971)	Skill	Rapport	Structure	Overload		+2
Aleamoni (1973, 1974)	Method of Instruction	Instructor		Course Content	General Course Attitude	+1
Aleamoni & Spencer (1973)	Method of Instruction (6%)	Instructor-Student Rapport (3%)		Course Content	General Course Attitude (30%)	+2 (47%)
Finkbeiner, Lathrop & Schuerger (1973)	Attitude Toward Method (19.1%)	Instructor-Student Rapport (20%)		Workload (7%)	General Course Attitude (26%)	+1 (50%)
Greenwood, Bridges, Ware & McLean (1973)	Obsolescence of Presentation	Rapport	Facilitation of Learning			+5
Kohlan (1973)		Rapport (24%)	Structure (44%)		Overall	+1 (77%)
McKeachie and Lin (1973)	Skill	(Group Interaction)	Structure	Difficulty		
Doyle & Whitely (1974)	Expositional Skills (26%)	Attitude Toward Students (26%)				+2 (99%)

**Numbers in parentheses are the percentage of variance accounted for.

TABLE 1B (continued)

Study	Skill	Rapport	Structure	Overload	General Course Attitude	Other
Jaeger & Freijo (1974)	5 Factors (Follows Hartley & Hogan, but factors do not agree here)	Student-Teacher Interaction (my label)			General Course Attitude (my label)	+5 (73%)
Pambookian (1974)	Skill	Rapport	Structure	Overload		+3
Pohlmann (1975)	Presentation of Material	Student Orientation	Assignments & Grading (?%)		Overall	+1 (77%)

This entire table(parts A and B) was reproduced from Bettencourt(1974)